Directed Evolution of Peptide Inhibitors of HIV-1 Entry

Directed Evolution of Peptide Inhibitors of HIV-1 Entry

A dissertation presented

by

Brian Donald Quinlan

to

The Division of Medical Sciences

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Virology

Harvard University

Cambridge, Massachusetts

October, 2013

Dissertation Advisor: Dr. Michael R. Farzan Author: Brian D. Quinlan

Directed Evolution of Peptide Inhibitors of HIV-1 Entry

Abstract

The conflict between HIV-1 and the host immune system plays out over a time-scale of months and years, and on a grander scale in the co-evolution of lentiviruses and the immune systems of their host species. Directed evolution of HIV-1 entry inhibitors using controlled randomization together with a display system offers a means of recapitulating one side of this conflict *in vitro* on an accelerated time-scale. To address limitations in existing display systems, we constructed a vector (pDQ1) integrating phage-display and mammalian-expression systems. This vector displays on phage when expressed in bacteria, and as an Fc-fusion when expressed in tissue culture, thus accelerating the iterative process of randomization, display, and characterization. We demonstrated the utility of this vector in the evolution of a CD4-mimetic peptide.

The HIV-1 envelope glycoprotein (Env) is the sole viral protein expressed on the surface of the virion, and it is subject to neutralization by the host antibody response. In adapting to this pressure Env has acquired inherent resistance to neutralization: Env has evolved such that its accessible epitopes are poorly conserved and its conserved epitopes are poorly accessible. Two epitopes are constrained by the need to interact with host proteins: the receptor- and coreceptor-binding sites. Access to these conserved sites is limited for molecules the size of antibodies. Targeting the binding sites with receptor-

mimetic peptides sidesteps accessibility issues in a way not possible for antibodies, while retaining access to the molecular toolbox of display technologies.

Using the pDQ1 vector, we adapted a natively-expressed version of a CD4-mimetic previously developed using non-natural amino acids. Development of a high-affinity natively-expressible peptide allowed for the fusion of this element to a tyrosine-sulfated CCR5 coreceptor-mimetic peptide to generate a double-mimetic peptide. Optimization of linker length and direction allowed for simultaneous binding of the double-mimetic to receptor- and coreceptor-binding sites. We characterized the CD4-mimetic and the double-mimetic peptides with binding, receptor-complementation and neutralization assays. Inclusion of a coreceptor-mimetic peptide improved avidity and eliminated CD4-like enhancement of viral infection. Our studies indicate that there is significant advantage to simultaneously targeting both conserved Env epitopes.

List of Abbreviations

293T	Human embryonic kidney cell line 293 expressing SV40 T-antigen
AID	Activation-induced deaminase
bNAb	broadly neutralizing antibody
C-terminus	Carboxy-terminus of a peptide or protein
CA	Capsid
CCR5	C-C chemokine receptor type 5
CD4	Cluster of differentiation 4
CD4-Fc	D1-D2 of CD4 fused to an IgG1 Fc domain
CD4-Ig	synonymous with CD4-Fc
CD4bs	CD4-binding site
CD4i Ab	CD4-induced antibody
CDR	Complement determining region in an antibody chain variable domain
CDRH1-3	CDRs 1 to 3 of a heavy chain variable region
CDRL1-3	CDR 1 to 3 of a light chain
CH1-3	Constant region Ig-domains 1 to 3 of an antibody light chain
CHAPSO	3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate
CHR	C-terminal heptad repeat, region of gp41
CL	Constant region Ig-domain of an antibody light chain
cryo-EM	cryo-electron microscopy
CXCR4	C-X-C chemokine receptor type 4
D1-D4	the four extracellular Ig domains of CD4
DMEM	Dulbecco's Modified Eagle Medium
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid, a divalent cation chelating agent
ELISA	Enzyme-linked immunosorbent assay
Env	Envelope glycoprotein
ER	Endoplasmic reticulum
ESCRT	Endosomal Sorting Complexes Required for Transport
Fab	Fragment antigen-binding, VL-CL linked to VH-CH1 by disulfide bond
Fc	Fragment crystallizable region, dimmer of hinge-CH2-CH3
FCS	Fetal Calf Serum
FP	Fusion peptide
gp120	glycoprotein 120, the surface unit of Env
gp160	glycoprotein 160, Env prior to cleavage into gp120 and gp41
gp41	glycoprotein 41, the transmembrane unit of Env
HDX	Hydrogen-deuterium exchange
HIV-1	Human Immunodeficiency Virus type 1
HR1	Heptad-repeat 1 (NHR)
HR2	Heptad-repeat 2 (CHR)
IC_{50}	Inhibitory concentration 50%
Ig	Immunoglobulin
IgG	Immunoglobulin gamma, isoform gamma antibody
IN	Integrase

ITC	Isothermal titration calorimetry
LEDGF/p75	Lens epithelium-derived growth factor
LTR	Long terminal repeat
MA	Matrix
MOPS	3-(N-morpholino)propanesulfonic acid, a Good's buffer
MPER	Membrane proximal ectodomain region of gp41
N-terminus	Amino-terminus of a peptide or protein
N1-N2	N-terminal domains of pIII that mediate entry into host bacterium
NAb	Neutralizing Antibody
NC	Nucleocapsid
NHR	N-terminal heptad repeat, region of gp41
ori	origin of replication
pIII	Minor coat protein of filamentous phage, product of gene 3
p6	HIV-1 *gag* gene product containing late domains
PBS	Phosphate buffered saline
PBS-T	PBS with the detergent Tween-20 (0.01-0.5%)
PDGFR	Platelet-derived growth factor receptor
PEG	Polyethylene glycol
Phe43	Phenylalanine 43 of CD4
PHI	Pre-hairpin intermediate
PIC	Pre-integration complex
POI	Protein of interest
PR	Protease, retroviral protease
R5 virus	CCR5 tropic HIV-1
RF	Replicative form of phage genome, which is double stranded
RRE	rev response element
RT	Reverse transcriptase
RTC	Reverse transcription complex
sCD4	soluble CD4, Usually D1-D2
scFv	single-chain variable fragment, VH and VL joined by flexible linker
SIV	Simian immunodeficiency virus
ssDNA	single-stranded DNA
STII	Heat-stable enterotoxin II
SV40	Simian vacuolating virus 40, Simian virus 40
TBS	Tris-buffered saline
TBS-T	TBS with the detergent Tween-20 (0.01-0.5%)
TBS-TC	TBS-T with the blocking agent sodium caseinate added
V1-V5	Variable regions 1 to 5 of HIV-1 Env
VH	Variable region Ig-domain of an antibody heavy chain
VL	Variable region Ig-domain of an antibody light chain
X4 virus	CXCR4 tropic HIV-1
X4R5 virus	Dual tropic HIV-1, CCR5/CXCR4 tropic HIV-1

Table of Contents

Abstract iii

List of Abbreviations v

Table of Contents vii

CHAPTER 1: INTRODUCTION 1

1. A. THE VIRAL LIFECYCLE 2

A Basic Overview of the HIV-1 Lifecycle *2*

1. B. THE ENVELOPE GLYCOPROTEIN 6

Env Glycoprotein Assembly and Structure *6*

The Consequences of Receptor Binding *8*

1. C. PEPTIDE AND PROTEIN INHIBITORS OF HIV-1 ENTRY 13

HIV-1 Neutralizing Antibodies *13*

bNAbs Recognizing the Receptor-Binding Sites *14*

bNAbs Recognizing Non-Binding Site Epitopes *19*

Peptide inhibitors of HIV-1 Entry *21*

1. D. PHAGE DISPLAY AND RELATED TECHNOLOGIES 24

Phage display and maturation of protein ligands *24*

Other Display Technologies *27*

Library construction techniques *30*

CHAPTER 2: Direct expression and validation of phage-selected peptide variants in mammalian cells 34

2. A. ABSTRACT 36

2. B. INTRODUCTION 36

2. C. METHODS 37

2. D. RESULTS 40

2. E. DISCUSSION 41

CHAPTER 3: Maturation of a CD4-mimetic peptide 42

3. A. ABSTRACT 44

3. B. METHODS 44

3. C. RESULTS 51
3. D. DISCUSSION 59

CHAPTER 4: A double-mimetic peptide efficiently neutralizes HIV-1 by bridging the CD4- and coreceptor-binding sites of gp120 61
4. A. ABSTRACT 63
4. B. INTRODUCTION 63
4. C. METHODS 64
4. D. RESULTS 68
4. E. DISCUSSION 74

CHAPTER 5: DISCUSSION 75
The dual-display vector technology *77*
An improved CD4-mimetic peptide *80*
The double-mimetic peptide *84*
Modeling escape pathways *89*
Closing Remarks *91*

List of Figures

Figure 1.1. The domain structure of core gp120 8

Figure 1.2. A model of HIV-1 entry 10

Figure 2.1. The design of pDQ1 39

Figure 3.1. Incorporation of mammalian cell expression and functional assays into the workflow of phage display 46

Figure 3.2. The generation of a soft-randomized library using pDQ1 47

Figure 3.3. The development of CD4mim6 53

Figure 3.4. Comparisons of CD4mim1-Ig and CD4mim6-Ig 55

Figure 3.5. A model of CD4mim variants with alterations indicated 56

Figure 3.6. CD4mim6, but not CD4mim6W, induces the CD4-bound conformation of the HIV-1 envelope glycoprotein 58

Figure 4.1. Receptor-mimetic constructs 64

Figure 4.2. Neutralization assays of single- and double-mimetic peptides 70

Figure 4.3. Infection of CCR5-positive, CD4-negative cells 71

Figure 4.4. Comparison of monomer and dimer forms of double-mimetic peptides 73

Figure 5.1. CCR5mim-Ig but not CD4-Ig binds two monomers of the HIV-1 envelope glycoprotein trimer 85

CHAPTER 1: INTRODUCTION

1. A. THE VIRAL LIFECYCLE

A Basic Overview of the HIV-1 Lifecycle

Viral entry. HIV-1 infection of a host cell begins with viral entry. The membranes of enveloped viruses, and the membranes of the cells they infect, are composed largely of lipids with negatively charged head groups. For viral entry to occur the virus must first overcome the tendency of these bilayers to repel one another [1]. The HIV-1 envelope glycoprotein (Env), the one viral protein exposed on the viral surface, must first attach to the cell by engaging in low affinity interactions with cellular surface molecules, which serve as attachment factors. Examples of these include the lectin DC-SIGN, an integrin ($\alpha_4\beta_7$integrin), as well as a number of proteoglycans such as heparan sulfate [2-4].

The first obligate binding event is to CD4, the primary receptor of HIV-1 [5, 6]. This high-affinity event initiates global rearrangements in the Env trimer spike, exposing the coreceptor-binding site, while bringing the virus in closer proximity to the plasma membrane [7-9]. Env then binds a coreceptor in a second obligate, high-affinity interaction. The coreceptor, usually a chemokine receptor (CCR5 or CXCR4 or both) determines the tropism of the virus (R5, X4, X4R5) and strongly influences the range of cells a given isolate may infect [10-14].

Receptor- and coreceptor-binding drive further rearrangements in the Env trimer that lead to insertion of the fusion peptide into the target membrane, resulting in fusion of

the viral and target cell membranes [15-17]. The consequence of this type I fusion process is the delivery of the viral core (capsid, RNA genome, and associated enzymes) into the cytoplasm of the cell.

Reverse transcription. The defining characteristics of retroviruses are the reverse transcription of a single-stranded RNA genome into a double-stranded DNA and the subsequent integration of that DNA into the host chromatin. Following delivery of the viral core to the cytoplasm are the closely coupled processes of reverse transcription and uncoating. The viral polymerase, reverse transcriptase (RT) starts to transcribe prior to uncoating, which is the disassembly of the conical capsid structure and the separation of the majority of capsid (CA) from the genome. But continuation and completion of reverse transcription requires uncoating and formation of the reverse transcriptase complex (RTC) [18, 19]. Reverse transcription begins by priming off of a Lys3-tRNA from a site immediately downstream of the 5' long terminal repeat (LTR). This produces a script that primes the 3' LTR to create a negative DNA transcript. The separate RNase H function of RT digests RNA from the RNA:DNA duplex. Priming of the positive strand is provided by an RNase H resistant element of the genome immediately upstream of the 3' LTR. The fragment protected by this undigested remnant primes on the negative strand 5' LTR, and is extended by the RT as DNA to create a positive strand LTR script, which then primes on the 5' LTR and is extended to give a complete positive strand [20, 21].

Nuclear import. At the end of reverse transcription, the 3' ends of the double stranded viral DNA are processed by integrase (IN). The RTC continues to lose CA as it matures into the pre-integration complex (PIC), which must reach the chromatin for

integration to occur. Retroviral PICs actively traffic to the nuclear periphery along the microtubules [22]. The PIC is larger than the nuclear pore, and is far too large for passive diffusion across the nuclear envelope. Many simpler retroviruses only efficiently infect dividing cells, using breakdown of the nuclear envelope during mitosis to access the host DNA [23]. However lentiviruses such as HIV-1 are able to efficiently infect non-dividing cells, and must therefore actively transport into the nucleus [24, 25]. The details of this process are not known, but interactions between CA and nuclear import factors are necessary for import to occur [26].

Integration. Inside the nucleus, integrase interacts with the host protein LEDGF/p75, which serves to tether the PIC to the host genome [27, 28]. This interaction influences target selection, with HIV-1 showing a strong preference for integration into active transcription units [29]. The recessed 3' hydroxyls, processed by IN in the cytoplasm, are then used by IN to attack the phosphodiester backbone at the site of integration in a process referred to as DNA strand transfer [30]. The final step in integration, gap repair, requires three enzymatic activities, most likely performed by host enzymes: a flap endonuclease such as flap endonuclease-1 must trim the 5'-overhangs of two mismatched bases; a polymerase is required to fill in the gap; and a host DNA ligase must close the single strand nick [31-33].

Early and late gene expression. Once integrated, transcription begins from the 5' LTR, which functions as a promoter [34]. The transcription products are spliced to produce mRNAs expressing the early genes *tat*, *rev,* and *nef*. Expression of later products depends on the accumulation of Tat and especially Rev [35, 36]. Rev binds to a unique regulatory sequence, the Rev Response Element (RRE) within the *env* gene, to promote

export of longer, unspliced, or partially spliced transcripts from the nucleus [37, 38]. Location of these transcripts in the cytosol allows for the expression of late genes, including the structural genes *gag, gag-pol,* and *env*. A full length, unspliced 9 kb primary transcript is packaged unmodified as the viral genome [39].

Assembly and budding. Virion assembly at the plasma membrane is mediated by the major structural protein gag, which brings together two copies of vRNA with its tRNALys3 primer, Env, and the enzymes RT, IN, and Protease (PR). Membrane bound gag and gag-pol assembles at a 20:1 ratio into immature spherical particles [40, 41]. The amino-terminal matrix (MA) domain is membrane bound via its myristoylation, which fixes the site of assembly at the plasma membrane and promotes Env association [42, 43]. A middle capsid (CA) domain mediates assembly and is the physical component of the mature core [44]. The nucleocapsid (NC) domain binds nucleic-acids via its zinc-finger motif, and directs incorporation of two copies of the genome into assembling virions [45]. The carboxy-terminal p6 region contains the so-called "late domain" which engages components of the ESCRT pathway critical to pinching off the membrane from the cell thereby completing the assembly process [46].

Due to a -1 frameshift at the end of *gag*, a fraction of the incorporated Gag is expressed as a longer product, Gag-Pol, containing the PR, RT, and IN domains. During and immediately following the budding event, PR cleaves the immature polyprotein into the protein products MA, CA, NC, p6, PR, RT, and IN. CA, separated from the membrane bound MA, reassembles into the conical shape of the mature core [47].

1. B. THE ENVELOPE GLYCOPROTEIN

Env Glycoprotein Assembly and Structure

Env trimer maturation. The HIV-1 Env, like all type I viral fusion proteins, has a trimeric quaternary structure. Env is synthesized as a single polypeptide, gp160, of approximately 850 amino acids, with an amino-terminal signal peptide that targets it co-translationally to the ER. In the ER, gp160 is heavily glycosylated, acquiring N-linked sugars on approximately 27 residues [48, 49]. Signal peptide cleavage is delayed, and passage of gp160 through the ER is slow, allowing for extensive isomerization of its 10 disulfide bonds [50-52]. Folded gp160 organizes into trimers and is transported into the Golgi apparatus, where it is cleaved by furin-like proteases into two fragments, the surface subunit gp120 and the transmembrane subunit gp41 [53-55], which remain non-covalently associated [56]. The surface gp120 subunit contains the receptor and coreceptor-binding sites, is highly glycosylated, and contains a number of highly variable regions [48, 57]. The transmembrane subunit, gp41, consists of a hydrophobic fusion peptide (FP) region, two heptad-repeat regions named NHR and CHR (also known as HR1 and HR2), a membrane proximal ectodomain region (MPER), a membrane-spanning domain, and a carboxyl terminal endodomain. Maturation of the Env by protease cleavage increases the conformational freedom of gp120, critical for fusion and immune evasion [58].

Ectodomain structure. The gp120 monomer is conventionally divided into five constant and five variable regions [59]. The constant regions of gp120 fold into a

conserved core structure, with the variable regions forming looped extensions from this core [57]. The conserved gp120 core consists of an inner domain that interacts with the gp41 stalk, a heavily glycosylated outer domain, and a bridging sheet minidomain that connects the two. The bridging sheet consists of four antiparallel β-strands, two from the V1/V2 stem in the inner domain and two from the outer domain. The bridging sheet and neighboring V3 stem contain residues of the coreceptor-binding site [57, 60-62].

Gp120 contains conformationally variant and invariant regions. The inner domain is structured as a 7-stranded β-sandwich from which emanate three loop 'excursions' and the N- and C-termini. (See Fig. 1.1) The β-sandwich, together with the N- and C-termini, form a gp41-interacting region. The loop excursions emanating from the β-sandwich form three stacked layers containing α-helical structure. The outer domain is an extension of the third topological layer. The outer layer and the inner β-sandwich are conformationally stable between liganded- and unliganded-states, with three inner domain layers being much more conformationally dynamic [63-65]. The topological layer model, based on crystal structures of the gp120 core bound to various ligands, is supported by hydrogen deuterium exchange (HDX) studies of bound and unbound gp120 core [66], as well as full length glycosylated gp120 [65]. Upon CD4 binding, the mobile inner domain layers all slow their rate of HDX, indicating a decrease in structural flexibility. The change is most dramatic for topological layer 2, in the bridging sheet residues, and in the helix immediately N-terminal to the sheet. For the outer domain, which has a net lower rate of exchange, decreases are seen only in the bridging sheet residues and in the CD4-binding loop [65].

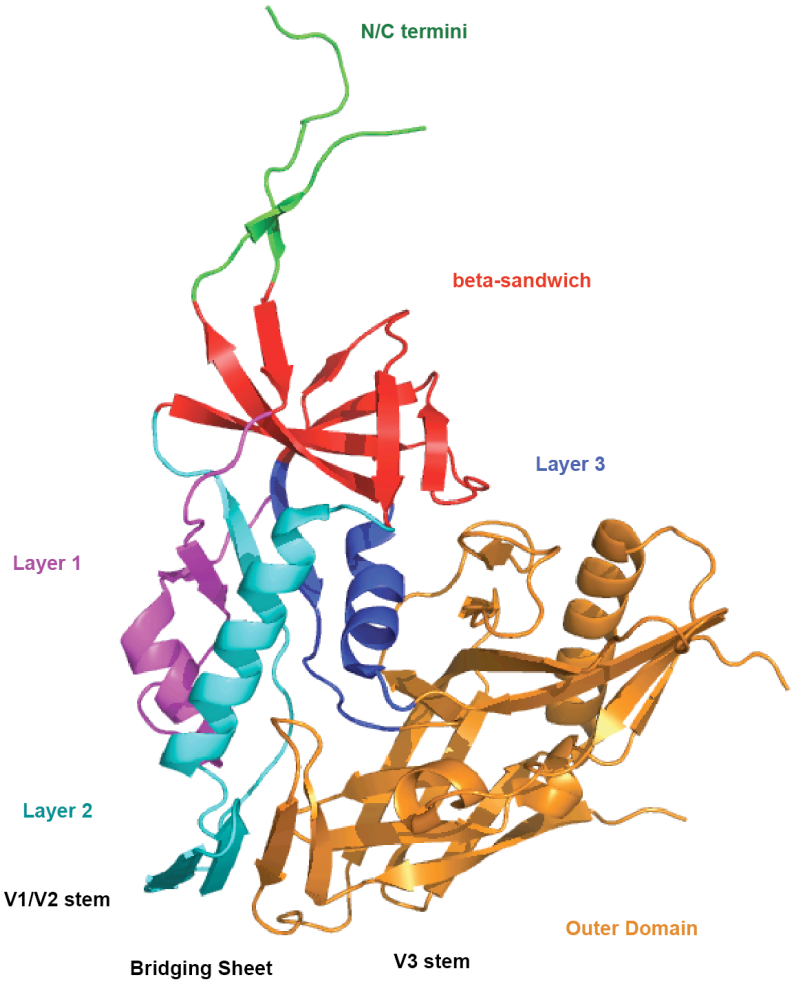

Figure 1.1. **The domain structure of core gp120.** Cartoon representation of CD4/48d-bound gp120 core with extended termini PDB: 3JWD after Pancera et al. [63]. The gp41 interacting domain is composed of N- and C-termini (green) and 7-stranded β-sandwich (red). The inner domain mobile layers 1-3 are represented in purple, cyan, and blue, respectively. The outer domain is represented in orange. Figure generated in PyMol.

The Consequences of Receptor Binding

CD4 binding induces changes in the gp120 tertiary structure. Entry begins with CD4 attachment to the CD4-binding site (CD4bs) on gp120 located within the cleft of the trimer. CD4 binds a large hydrophobic cavity in gp120 between the inner, outer and bridging sheet domains, with the cavity being capped and partially occupied by a

critical phenylalanine (Phe43) of CD4 [67, 68]. Comparison of liganded and unliganded SIV gp120 structures initially indicated that the bridging sheet structure was only formed upon CD4 binding [68]. Furthermore, the crystal structures of core gp120 bound to several antibodies revealed conformations in which the bridging sheet was not formed [69, 70]. However, these observations are not in good agreement with the available cryo-EM data for trimeric Env [71, 72]. Moreover, an "extended core" of gp120, containing more of the V3 and an intact N terminus, has now been crystallized in unliganded and liganded states with the unliganded extended core forming a structure very similar to that of the CD4-liganded state. Biochemical analyses favor a model in which the CD4-liganded state is a ground state, and the unliganded gp120 is discouraged from sampling this state by the V1/V2 and V3 regions [73].

CD4 binding induces changes in the Env trimer quaternary structure. In addition to the changes in tertiary structure of the gp120 monomer induced by CD4 binding, there are distinct changes in the quaternary structure of the Env trimer that expose the coreceptor-binding site. Cryo-electron tomography has been used to visualize the Env trimers in liganded and unliganded states. CD4 binding corresponds to a rotation of the monomer away from the central axis of the trimer. This rotation moves the trimer from a "closed" to an "open" conformation in which the V3 stem and associated coreceptor-binding site are located at the apex of the spike, making them accessible for binding to a coreceptor in the target membrane [71, 74].

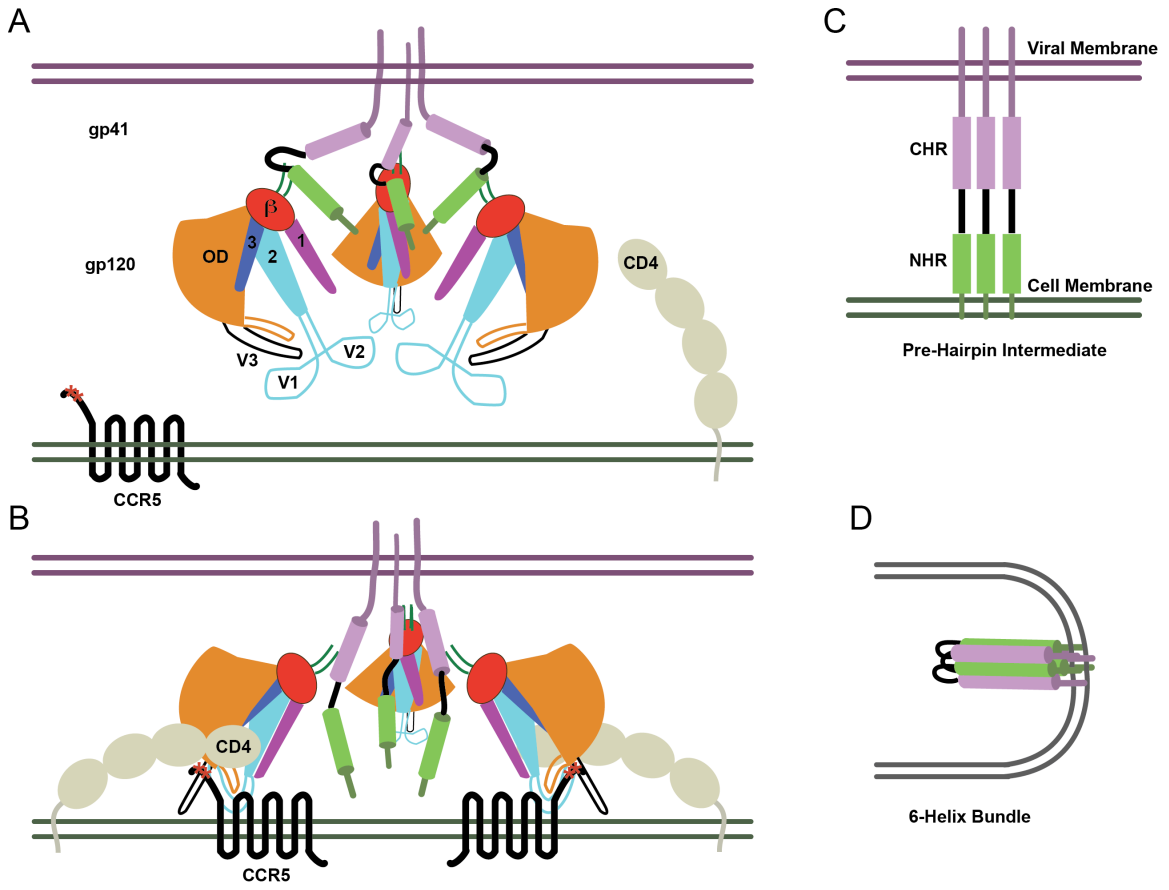

Figure 1.2. A model of HIV-1 entry. gp120 subunits are represented with schema of topological layers. The outer domain is represented in orange. The inner domain layers one, two, and three shown in purple, cyan, and blue, respectively. The invariant β-sandwich (red) is shown with N- and C-termini (green) emerging at the interface with gp41. The gp41 subunit is shown with fusion peptide as a dark green line attached to the N-terminal Helix Region (NHR) in light green, the C-terminal Helix Region (CHR) in light purple, and the transmembrane and endodomain are represented as a dark purple line. CD4 is shown with its four Immunoglobulin domains in gray, and the seven transmembrane domain receptor CCR5 is represented in black, with red stars representing site of tyrosine sulfations. **A,** Prior to CD4 binding, V1/V2 (light blue loops) and V3 (black loop) hold Env in a closed conformation. **B,** After binding CD4 and CCR5, the bridging sheet forms between the V1/V2 stem and beta-20-21 hairpin (orange line) of the outer domain. CCR5 binds at the bridging sheet with tyrosine sulfated N-terminus binding at the base of the V3. **C,** Following receptor and co-receptor binding, the gp41 fusion peptide is inserted into the target cell membrane forming the Pre-Hairpin Intermediate (PHI). D, The viral and cellular membranes fuse, and the six-helix bundle forms with the CHRs packs into grooves formed by the interior bundle of NHRs.

CD4 binding to gp120 is an entropically unfavorable process. CD4-gp120 binding is notable as an extreme example of an enthalpy driven process. There is a high entropy cost to CD4-gp120 binding that is compensated for by gains in enthalpy [75]. The large negative entropy change appears to be due to loss of sampling of alternative structures rather than folding of unstructured polypeptide [66, 73]. CD4 binding imposes direct constraints on layers 2, 3 and the outer domain, fixing their orientation and decreasing conformational mobility. The consequence of the high entropic barrier is a "conformational masking" of the CD4-binding site epitope [67]. Thus conserved conformations are masked from antibody recognition by the high energy barrier, ensuring that the antibodies that emerge early in infection will be to non-conserved epitopes, which do not require the fixing of mobile topological layers for recognition.

CD4 binding exposes the coreceptor-binding site. The major HIV-1 coreceptors are the chemokine receptors CCR5 and CXCR4. Both are seven-pass integral membrane proteins with long free N-termini containing sulfated tyrosines [76, 77]. Binding of R5 tropic isolates in particular is dependant on the strongly electronegative sulfotyrosines in the N-terminus, which bind to pockets at the base of the V3 loop [62]. Receptor and coreceptor binding to the Env glycoprotein are not strictly serial, as they are mutually reinforcing and cooperative events [75, 78]. Binding of CD4 or of the coreceptor proxy 17b Ab leads to an opening of the Env trimer conformation [74]. Access to the coreceptor-binding site is limited by V1/V2 [79, 80], and elimination of these regions confers a degree of CD4 independence to Env [81]. Masking of the coreceptor-binding site prior to CD4 engagement serves to limit exposure to neutralizing antibodies until the target membrane is close enough to limit access to this site [82].

Changes in gp41 due to CD4 binding. The conformational changes that take place in gp120 upon receptor and coreceptor binding lessen the association with gp41, and in some lab adapted strains the ectodomain is "shed" when exposed to soluble CD4 [83]. At the same time, structural rearrangements in gp41 are initiated, leading ultimately to membrane fusion by a mechanism conserved among all type I viral fusion proteins [16, 17]. The N-terminus of gp41 contains a hydrophobic fusion peptide, which is sequestered in the center of the trimer. Engagement with receptor and coreceptor leads to extension of the N-terminal fusion peptide away from the virus and into the target cellular membrane. This extended structure is referred to as the pre-hairpin intermediate (PHI), and this state persists long enough to be a target of fusion inhibitors [84-87].

Formation of the six-helix bundle. The energy that drives viral fusion is generated by the rearrangement of the metastable PHI into a six-helix bundle. Between the fusion peptide, now in the cellular membrane, and the transmembrane of gp41, there are two regions of repeating amphiphilic heptads. The amino-terminal heptad repeat (NHR) is next to the fusion peptide, and the carboxy-terminal heptad repeat (CHR) is near the transmembrane region, with the two regions linked by a disulfide containing "immunodominant region", so called because of its high immunogenicity. Arranged in a trimer, the NHRs form a bundle with three hydrophobic grooves. When the PHI collapses into the six-helix bundle, the CHR fill these grooves, a conformation described as the trimer of hairpins [15, 88]. The formation of this highly stable structure brings the viral and plasma membrane into close proximity, promoting lipid mixing and viral fusion [16, 89].

1. C. PEPTIDE AND PROTEIN INHIBITORS OF HIV-1 ENTRY

HIV-1 Neutralizing Antibodies

The Three Faces of Env. A lentiviral infection can persist for years before killing its host, requiring the virus to be exposed to the host immune system for a long period of time. The complexity and variability of the HIV-1 envelope glycoprotein is driven in part by the need to evade immune recognition. In the early stages of infection, the first Env specific antibodies to appear are non-neutralizing antibodies recognizing gp41 [90]. Non-neutralizing antibodies that recognize gp120 emerge over the next few weeks, with neutralizing antibodies (NAbs) emerging months after primary infection. These early emergent NAbs are narrowly neutralizing, and do not recognize heterologous viruses [91-93]. Such autologous neutralizing antibodies are raised against regions of high variability in gp120, such as the V1/V2, and the V3 regions [92, 94, 95]. Broadly neutralizing antibodies (bNAbs) emerge after years of persistent infection [96-98].

A large number of Env antibodies have been isolated from HIV-1 patient B-cell repertoires, have been characterized, and have served as tools for probing Env structure and function. Prior to the availability of a crystal structure, gp120 was mapped using gp120 mutants together with antibody binding and competition assays. Antibodies against gp120 were classified by whether they competed with sCD4 for binding, whether their epitopes were inducible, whether they neutralized, and the location of the mutations that abrogated their binding [99]. These techniques allowed for the mapping of neutralizing and non-neutralizing antibodies to theoretical surfaces of gp120 [100]. Neutralizing and non-neutralizing faces were identified, distinguishing between epitopes

that are exposed or concealed in the intact trimer [101]. With the availability of a gp120 core crystal structure, a third face was identified. This is termed the "silent" face, due to the scarcity of antibody epitopes in this region, and it corresponds to that part of gp120 concealed beneath the so-called glycan shield [57, 60]. The subset of antibodies that are broadly neutralizing (bNAbs) has been especially well studied. They are potentially useful as inhibitors in their own right, and represent the subset that one would want to elicit with a vaccine. In addition, they highlight the vulnerabilities of HIV-1 helpful in the design of inhibitors of HIV-1 entry.

bNAbs Recognizing the Receptor-Binding Sites

CD4-binding site antibodies. For antibodies to be neutralizing, they must have access to their epitopes, either in the unliganded state, or transiently during the entry process. For an antibody to be broadly neutralizing, its epitope should be conserved. Among the most conserved epitopes are those of the receptor-binding sites. Functional requirements restrict the variability of the CD4-binding site (CD4bs), and ensure it is accessible. HIV-1 limits neutralization at the CD4bs through steric hindrance and conformational masking. The site is recessed in the trimer cleft and is protected by a glycan shield, thus limiting access to the site by molecules larger than the first immunoglobulin domain of CD4. Conformational masking refers to the high entropy barrier to fixation of this epitope [67].

The first well-characterized CD4bs antibody was b12 [102], which has a neutralization breadth of approximately 40-50% [103, 104]. Neutralization breadth is

commonly defined as the percentage of Env isolates that are neutralized to a 50% inhibitory concentration (IC_{50}) by a concentration of 50 μg/mL across a wide array of isolates with the results obtained being dependant on the composition of the panel used. The b12 antibody binds with a lower entropic penalty than CD4 but also with a lower enthalpy gain [67]. Binding of CD4 imposes constraints on layers 2, 3 and the outer domain. In contrast, b12 binding constrains only layer 3 and the outer domain, leaving layer 2 of the inner domain, which contains the V1/V2 stem, unconstrained [63, 70]. The CDRL3 of b12 places a tyrosine in the Phe43 cavity of the CD4-binding site, while the CDRH3 contacts with the beta20-beta21 strands that are the outer domain contributions to the bridging sheet. Thus b12 overlaps with the coreceptor-binding site, without stabilizing the bridging sheet contact between layer 2 and the outer domain. With fewer constraints placed on binding, one would expect that a greater number of conformational states would be able to bind b12. This has been probed using cross-linking studies. Glutaraldehyde cross-linking was used to freeze the envelope glycoprotein in a mixture of states, and ligands were evaluated for binding to cross-linked and untreated gp120. Relative binding ratios of cross-linked to untreated gp120 were presumed to reflect the proportional occupancy of states that the ligand can bind. From these studies it was observed that sCD4 binding was much more restricted by cross-linking than was b12 [73]. Perhaps as a consequence of the divergence of the b12 epitope from the CD4bs, this antibody as mentioned lacks breadth. This lower breadth suggests that the barrier to escape from this antibody is low. Some of the limitations on the breadth of b12 may also be a consequence of quaternary limitations. Residues that impact gp120 binding have been identified, but these do not directly correlate with neutralization. Rather isolates

known to have a more open conformation in the trimer are typically more neutralization sensitive [105]. Binding of b12 to trimer requires an induction of a quaternary shift that may reflect steric restraints on its binding [71, 74, 106].

A number of much more broadly neutralizing (~90% of isolates tested to IC_{50} at 50 μg/mL) CD4bs antibodies have recently been identified, such as VRC01, 3BNC117, and NIH45-46 [103, 107]. The best studied of these is VRC01, which has a neutralization breadth of 91% compared to 41% for b12 in the same study. VRC01 binds residues that closely match CD4 contacts. Interactions are almost entirely mediated by the heavy chain, and VRC01 light chain contains an unusual deletion of two amino acids in the CDRL1 that reduces a steric clash with Loop D of gp120 [108]. VRC01 binding to monomeric gp120 has a similar thermodynamic profile to that of CD4/gp120 binding, with high entropic cost and a high enthalpic pay-off. As with CD4/gp120 binding, the entropic penalty of VRC01/gp120 binding partially lowers the entropic barrier to binding for coreceptor-binding site antibody 17b, and these antibodies bind cooperatively to gp120 [103, 109, 110]. However, VRC01 is not a perfect proxy for CD4. In cross-linking experiments VRC01 behaves more like b12, binding a larger proportion of locked conformers than sCD4 [73]. Cryo-EM studies of the Env trimer bound to VRC01 and 17b shows an interesting pattern of quaternary structures. When mixtures of VRC01 and 17b are bound to trimer, two populations are observed: a closed conformation with VRC01 bound, resembling the unliganded trimer, and an open conformation with bound 17b, resembling the CD4-liganded trimer. The suggested model is that when VRC01 binds to Env, it locks in a closed conformation which prevents 17b binding, but if 17b is prebound, locking in the open conformation, VRC01 will still bind that open

conformation [74]. The conformational promiscuity of VRC01 binding suggests a partial explanation for its great breadth.

CD4 induced antibodies. Like the CD4-binding site, the coreceptor-binding site, located at the bridging sheet and the V3 base, is constrained by functional requirements, and residues within this site are among the most conserved on gp120 [111]. Unlike the CD4-binding site, the coreceptor-binding site does not need to be accessible in the unliganded state, and access to this site is limited prior to CD4 binding. Antibodies that bind CD4 induced epitopes in this region (CD4i antibodies) are believed to bind like the coreceptor, and their binding has been used to make inferences about the nature of the structural transitions that take place during entry. The prototypical CD4i antibody is 17b, which binds the bridging sheet. Like CD4-binding, the binding of 17b to gp120 incurs a high entropic penalty; however, when CD4 is prebound, the entropic penalty of binding 17b has already been paid [109]. Consideration of the 17b binding surface on gp120 provides insight into this entropic barrier: 17b binding to the bridging sheet fixes inner domain layer 2 to the outer domain, thus restraining inner domain layer 3 and locking gp120 in the CD4-liganded conformation [57].

Access to the coreceptor-binding site is restricted by the V1/V2. Deletion of the V1/V2 region allows for binding of CD4i antibodies to the gp120 monomer and can render HIV-1 somewhat CD4 independent [81]. Additionally, there are steric restrictions on access to the coreceptor-binding site in the context of the trimer. CD4i antibodies can bind to unliganded trimer in lab-adapted isolates, which are known to assume an open conformation, but are limited in their ability to bind primary isolates [82]. By cryo-EM, 17b is only observed bound to the open conformation [74]. CD4i antibodies are poor

neutralizers of primary isolates despite their ability to bind sCD4-induced trimers. Inefficient neutralization by these antibodies may be attributed to the temporal and steric restrictions on access to their epitope. The CD4i epitope is unmasked only in the window of time between receptor binding and coreceptor binding. However, at this time, the virus is in close proximity to the cellular membrane, which sterically limits antibody access. Consistent with this, and unusually for HIV-1 neutralizing antibodies, the monovalent Fab forms of these antibodies are more active than the larger divalent IgG forms. Antibody addition experiments show that this difference is attributable to their differing ability to bind Env after receptor engagement [82]. There may therefore be less selective pressure at the coreceptor-binding site, explaining its conservation, and suggesting it would make a good target for inhibitors small enough to access this site.

The N-terminus of the HIV-1 coreceptor CCR5 contains several sulfotyrosines critical to Env binding, and the ability to bind sulfotyrosines is retained even in X4 tropic viruses [78, 112]. A number of CD4i antibodies contain long flexible CDRH3s rich in acidic residues and tyrosines, which promotes their sulfation [113]. These antibodies bind the sulfotyrosine binding pockets at the base of the V3 loop, which is part of the coreceptor-binding site [62, 114, 115]. The CDRH3 of 412d contains two sulfated tyrosines, and has been crystallized bound to gp120. The gp120/412d co-crystal reveals two sulfotyrosine binding pockets at the base of the V3 and shows that, as with 17b, the bridging sheet is fixed [62]. A similar pattern with glutaraldehyde fixation is observed for 412d as for 17b, suggesting that the conformational masking evident in the isothermal titration calorimetry (ITC) measurements of 17b binding to the bridging sheet would hold for binding to the sulfotyrosine binding pockets [73]. A related antibody E51, from the

same VDJ recombination event in the same patient, has an even longer CDRH3, and like the N-terminus of CCR5, contains four potentially sulfated tyrosines [112, 114]. It is this CDRH3 that is the starting point for the CCR5 mimetic peptide that participates in the double mimetic we generate and characterize in Chapter 4.

bNAbs Recognizing Non-Binding Site Epitopes

Antibodies recognizing gp120 glycosylations. Aside from antibodies recognizing the receptor- and coreceptor-binding sites, there are several classes of bNAbs targeting other Env glycoprotein epitopes [116]. N-glycosylations can form and contribute to these additional epitopes, partially and wholly [117]. It is somewhat counterintuitive that one would find bNAbs dependent on glycosylations: protein-glycan interactions are relatively weak, anti-glycosylation antibodies tend to be autoreactive, and most glycosylation sites are poorly conserved [116, 118]. However, some of these glycosylation sites can be relatively conserved due to their role in protein folding, and these conserved glycans participate in a number of bNAbs.

Moreover, the density of glycosylation sites on gp120 is itself a distinctive feature, and this density results in less efficient trimming of mannose residues during processing in the ER and Golgi [119, 120]. An epitope containing multiple such residues can be recognized as non-self. A well-characterized antibody of this type is 2G12, which recognizes high mannose type glycosylations in the conserved regions at the base of the V3 loop, and within the V4 loop [99, 121]. 2G12 has a unique domain swap structure, where the VH domains of the two Fabs of the IgG dimer cross-pair with the opposite VL

domains. This unique domain structure brings the binding sites of the two Fab arms into close proximity, allowing for recognition of the multivalent epitope provided by the clustered glycosylations [122]. Although the 2G12 epitope is not itself critical to Env function, this antibody appears to neutralize by blocking access to the coreceptor-binding site [123].

A different set of N-glycan dependent antibodies is that represented by the bNAbs PG9 and PG16. These antibodies reach through the glycan shield to recognize an epitope made of two N-glycosylations and a stretch of peptide in the V1/V2 region of gp120 [124, 125]. The epitope for PG9 contains elements from more than one unit in a trimer, giving this antibody the unusual characteristic of efficiently recognizing native trimer while poorly binding the gp120 monomer [126]. Binding of a single antibody to one monomer of a trimer blocks CD4 binding to all monomers [127].

MPER antibodies. Unlike gp120, gp41 is relatively inaccessible, and anti-gp41 antibodies are mostly non-neutralizing. An exception to this rule is the class of antibodies recognizing the membrane-proximal external region (MPER), an induced epitope that, for many isolates, is only accessible during entry for many isolates [128, 129]. The MPER is a stretch of gp41 that inflects to rest along the lipid surface before traversing the membrane; a structure that has been described as the "knee" of gp41. Two of the anti MPER antibodies so far identified, 4E10 and 2F5, have long hydrophobic CDRH3s that bind phospholipids, which may increase their concentration at the membrane during the exposure of this epitope [130, 131]. A third such antibody, 10E8, recognizes an epitope shifted several residues from the membrane, and does not bind lipids [132]. MPER antibodies do not appear to account for a major component of the

neutralizing antibodies elicited in the course of infection [116]. These antibodies tend to be autoreactive, which may account for their scarcity [133, 134].

Peptide inhibitors of HIV-1 Entry

Peptide inhibitors of HIV-1 replication. The exposure of the envelope glycoprotein to the extracellular milieu makes it vulnerable not only to antibodies, but also to peptide inhibitors. Peptide inhibitors, like antibodies, face the same set of challenges in neutralizing a highly variable target. Although compared with antibodies peptides suffer from poor bioavailability, they are synthesizable and can potentially reach targets that are inaccessible to antibodies. Like antibodies, peptides can be expressed in biological systems, and improved through display technologies.

A CD4-mimetic peptide constructed by loop grafting. The CD4-binding site on gp120 contains a hydrophobic pocket critical to binding. This pocket fits the Phe43 residue on CD4, which resides on a highly surface exposed β-turn [57]. This end of this β-turn in CD4, contained in residues 40-48, is responsible for 63% of the contact area. The Phe43 alone provides 23% of this contact area [57]. However, linear peptides from this region fail to inhibit HIV-1 entry, presumably due to a lack of secondary structure, and the extra entropy cost associated with folding a disordered peptide [135]. Attempts to enforce a β-hairpin through disulfide cyclization [136], or in a synthetic cyclized peptidomimetic failed to produce significant specific binding activity [137]. The laboratory of Claudio Vita solved this problem by grafting the residues 40-43 (QGSF) onto the β-hairpin of a scorpion-toxin scaffold [138].

Scorpion toxin consists of an N-terminal α-helix stapled to a C-terminal β-turn by three disulfide bonds. This fold tolerates substitutions in the α-helix or the β-turn, and even the deletion of a disulfide bond [139-141]. A large number of these toxins are observed in nature, and there is no sequence conservation other than the disulfide bonds [142]. This scaffold has been used to present α-helix derived from an anti-fibrinogen antibody [139], and to engineer a metal-binding motif in the β-turn of the scaffold [140].

Following the successful grafting of the Phe43 loop onto the scaffold, a combinatorial approach was used to optimize and obtain a high affinity, semi-synthetic peptide using non-natural amino acids [143]. Significantly this peptide was shown to induce the CD4i epitope of the coreceptor-binding site antibody 17b [144, 145]. This peptide has been co-crystallized with gp120 and 17b, and its binding induces the same conformation as does CD4 [57, 146]. This is in contrast to the structures of the CD4bs antibodies b12, b13, and F105 [69, 70].

Coreceptor-mimetic peptides. High affinity binding of HIV-1 Env to the coreceptor CCR5 is dependant on the sulfation of CCR5s N-terminal tyrosines, and a number of antibodies to the coreceptor-binding site on gp120 contain sulfated tyrosines in their CDRH3s [114]. Sulfated peptides based on the N-terminus of CCR5 or on the CDRH3s of CD4i antibodies can be used to inhibit HIV-1 entry [112, 147]. The sequences of the CCR5 N-terminus and the CDRH3s of E51 and 412d may be aligned by their tyrosine patterns; the 412d peptide can be swapped in for the N-terminus of CCR5 to generate a coreceptor capable of supporting infection by R5 tropic virus [115]. The strength of this interaction derives from the negatively charged sulfates interacting with

positive residues in the binding pockets at the base of the V3, while aliphatic residues that line the pockets interact with the aromatic ring of the tyrosines [62]. The strength of these electrostatic contacts is sufficient such that the exact secondary structure of the peptide backbone is less constrained. The crystal of the unbound E51 does not resolve discernable structure in the CDRH3, implying this loop is relatively unconstrained. As such, the free peptides derived from the CDRH3 or the N-terminal fragment from CCR5 can bind without a scaffold anchor. Indeed, a short highly sulfated peptide derived from the CDRH3 of CD4i antibody E51, and improved through mutagenesis, is the strongest coreceptor mimetic yet identified [112, 148].

Fusion inhibitors. Following receptor and coreceptor binding, gp41 extends and inserts its N-terminal fusion peptide into the target membrane, forming a high-energy state pre-hairpin intermediate. Subsequently this intermediate collapses into a six-helix bundle, driving fusion of the two membranes. This six-helix bundle is formed by two heptad repeat sequences in the N- and C-terminal regions of the gp41 ectodomain (NHR and CHR) [149]. In the six-helix bundle, the NHR segments from three gp41 ectodomains form a central trimeric coiled-coil, defining three hydrophobic grooves. Folding over to form a hairpin, the CHR helices pack into these grooves in an antiparallel fashion [15, 88, 150]. In the prehairpin extended conformation, the NHR and CHR segments are not yet associated and are transiently accessible to fusion inhibitors. Inhibitors based on segments of the CHR are called C-peptides and work by packing into the central hydrophobic grooves in place of the CHR [87, 151]. One such peptide, T20 (Fuzeon, enfuvirtide) has been approved for use in HIV-1 infected individuals as a salvage therapy. Escape mutants to this peptide readily emerged, some of which work by

increasing the kinetics of fusion to decrease the availability of the pre-hairpin intermediate target [152].

1. D. PHAGE DISPLAY AND RELATED TECHNOLOGIES

Phage display and maturation of protein ligands

The phage display technique. Phage display of proteins is a conceptually simple, yet powerful technique to connect a protein of interest (POI) to the genomic material that encodes it. This allows for the selection of genomic material based on the binding properties of its product. Filamentous phage were among the first commonly used cloning vehicles [153-155], and some of the features that made them attractive tools for molecular cloning also make them attractive as a display technology. The genome is small (6.4 kb), yet is able to tolerate insertions into the major intergenic region, which can more than double its size. It is non-lysogenic, greatly simplifying its handling, and allowing it to persist in culture. Both of these qualities derive from the processive extrusion process by which new virions are produced. Once packaging has begun at the morphogenic (packaging) signal, the phage will continue to ratchet the genome through the inner membrane, assemble in periplasm, and extrude the nascent filament through a pore in the outer-membrane until the end of the genome is reached. This obviates the need for membrane lysis and also means that the capsid contains no intrinsic determinants of length or content [156].

Polyvalent Display. The pIII minor coat protein of filamentous phage serves two distinct functions in the viral life cycle. The C-terminal domain, connected to the tip of the phage, is responsible for liberating the phage from the host at the end of the assembly process. The two N-terminal domains (N1 and N2) mediate receptor binding and entry into a target cell. N2 engages the primary receptor at the tip of the pilus. The pilus retracts bringing N1 into proximity with the coreceptor, the membrane protein tolA, to initiate entry of the genome into the target cell [157]. Importantly, these domains are modular, connected to each other by long flexible linkers of exceptional regularity ($[EG_3S]_3EG_3$ and $[G_3S]_3[EG_3S]_4G_3SGSG$) and have been likened to beads on a string. Phage display with replication competent phage (e.g. fd-tet) is achieved by merely adding another bead to the N-terminus of the string [158, 159]. However, the large size of these vectors, and the fact that the inserted protein can interfere with replication, limits the utility of replication competent systems [160, 161]. Furthermore, replication competent vectors are inherently polyvalent: all of the 3-5 copies of the minor coat protein will be fusions to the protein of interest, which introduces the potential for multipoint attachment and undesirable avidity effects. These problems were addressed by the development of phagemid display vectors [162].

Phagemid vectors. A phagemid is a pBR322-derived plasmid to which a phage origin of replication has been added. The phage ori is the start site of single strand synthesis and the morphogenic (packaging) signal. These elements are sufficient to ensure that any circular DNA construct containing the ori in the context of replicating phage will be packaged as single-stranded DNA (ssDNA). The combination of a phage ori with an unconstrained pUC origin was found to generate dominant defective

interfering particles. In the host bacterium, filamentous phage go through a double-stranded stage, the replicative form (RF). The unconstrained plasmid origin causes this form to be amplified, which then ensures that the majority of the circular ssDNA scripts generated upon rescue with helper phage will be phagemid in origin.

This interference is caused by the amplification of the double stranded RF provided by the plasmid origin, which ensures that the majority of circular ssDNA forms available for packaging were from that source, swamping the genomes of replication competent phage [156]. This has been widely used to generate large quantities of ssDNA for molecular biological applications such as Kunkel mutagenesis.

Monovalent display. A phagemid display vector also encodes a coat protein fused to the POI. Since the replication competent helper phage supplies a fully functional copy of pIII, the protein of interest can replace the N-terminal domains of the phagemid encoded pIII, with the C-terminal domain ensuring incorporation into phage particles [162]. The valency of display is then controlled by use of an inducible Lac promoter. In the absence of induction, low expression of the POI-pIII fusion ensures that most particles will have only the wild type pIII and a small minority will have only one copy of the fusion. As expression increases, more POI-pIII fusion will be expressed, and more particles will have multiple copies of the POI-pIII fusion. At least one copy of wild type pIII must present for the particles to be infectious [162, 163].

Phage panning. Panning is the selective enrichment of phage particles with the desired binding characteristics. Panning can be solution-phase or solid-phase. The most basic form of solid phase panning involves adherence of the bait protein to the plastic

surface of an ELISA plate. More typically, beads are used as a substrate. Phage particles are allowed to bind to the bait, and then are removed in a series of washes. Stringency is introduced by the wash steps, which select for a slower off-rate. With solution-phase panning, phage and bait are allowed to encounter in solution, and then are coprecipitated with beads. In both cases, following the wash steps, phage are eluted from the substrate and amplified, a process that may be repeated several times until the library has been reduced to an appropriate size for evaluation.

Other Display Technologies

Surface display techniques. Though phage display is by far the most common surface-display technique, there are a number of other technologies used for display, each with their own advantages and disadvantages. Filamentous phage display has a number of weaknesses which competing techniques seek to address. For example, some proteins will not express in the periplasm of *E. coli* or will not express as a fusion to pIII. Also, phage particles are too small to use cell-sorting techniques, which could otherwise be used with labeled soluble bait to control for unpredictable valency and undesirable avidity effects. Moreover, some proteins require post-translational modifications, in which case no bacterial system will suffice.

Each alternative system has its own tradeoffs. The parameters that define the utility of these display systems are the complexity of the starting library, the number of unique library members that can be evaluated, and the relative ability to distinguish between binders of differing affinities. Different systems will also have different

intrinsic limits, such as on the size, hydrophobicity, and modifications of the proteins that can be displayed. Artifacts introduced at the amplification step caused by factors such as codon usage, metabolic burden, and toxicity, can confound discrimination between library members in ways that differ between systems.

Bacterial display. Use of whole bacterial cells as a display platform allows for a great number of protein scaffolds to be used in contrast to phage display which is limited to fusion with the phage coat proteins [164]. The limits on library size are similar to those for phage display and are dependant on transformation efficiency. Evaluation of library members is by cell-sorting, and the size of the library that can be evaluated is subject to the throughput limits of this technique, typically in the range of 10^7 cells [165]. Magnetic sorting, with bait protein conjugated to magnetic beads, can be used to consider a library of arbitrary size. However here avidity is a factor so this approach is typically used to narrow the size of the library in the initial rounds of selection. Fine distinctions between binders are distinguished through later cell-sorting rounds with limiting amounts of fluorescently labeled bait [164-166].

Yeast Display. Display of peptides and proteins on the surface of *Saccharomyces cerevisiae* has been achieved by fusing the POI to the C-terminus of the Aga2p mating adhesion receptor. Yeast display allows for the presentation of large proteins, and of those that can only fold in a eukaryotic system. Library complexity is regularly in the 10^7 range and has been reported to be as high as 10^9 members [167]. Many thousands of copies of the POI may be displayed on the surface and as with bacterial display, selection is typically achieved by cell-sorting on a fluorescently labeled target protein, with similar throughput limits to selection in the range of 10^6-10^7 individual cells [168], and

consideration of larger libraries is accomplished by first selecting with magnetic beads before rounds of selection by cell-sorting [167, 169, 170]. An additional advantage for library construction for the display of a heterodimeric protein, such as Fabs, is that mating can be used to achieve large combinatorial libraries [171, 172]. A limitation of yeast display is post-translational modifications are significantly different between yeast and mammalian cells.

Mammalian surface display. Display of proteins on the surface of mammalian cells is a technique analogous to yeast display that allows for native mammalian folding and post-translational modifications. The basic requirements of a mammalian surface display vector are a POI expressed as a fusion to a transmembrane domain for cell surface presentation, and a means of maintaining POI expression. One way of achieving the latter is by transfection of a vector containing an episomal origin of replication and a selectable marker [173, 174]. An example of such a system is the commercially available pDisplay vector, which utilizes a PDGFR transmembrane domain, a neomycin-selectable marker as well as an SV40 origin for episomal maintenance [175]. Cell number in mammalian tissue culture limits library size, typically in the range of 10^6 cells [176].

An alternative system, and one with a means of approximating a single POI per cell, is expression of a POI-transmembrane fusion from a lentiviral vector. Co-transfection with lentiviral packaging plasmids and an envelope protein is used to generate pseudovirus particles, which can then be used at a low multiplicity of infection to generate the cell library. From this point on the system is treated the same as an episomal system and is subjected to successive rounds of passage followed by selection [177].

As with yeast display, in mammalian cell-surface display selection is typically achieved with cell-sorting on a fluorescently labeled soluble target, so that target avidity is not an issue. Here again, initial rounds of selection may be performed with magnetic beads to consider an arbitrarily large library size. An interesting feature of mammalian display is the possibility of utilizing the somatic hypermutation pathway to introduce diversity into antibody libraries. Exogenous activation-induced cytidine deaminase (AID) enzyme can be expressed introduced into standard cell lines [174, 178], or a lentiviral system can be used to transduce hypermutating B-cell lines [177, 179]. Thus at any point in the iterative improvement of an antibody, a previously selected libraries can be rediversified.

Library construction techniques

Antibody repertoires. Phage display has been widely used as a means of searching through vast antibody repertoires. Sets of light and heavy chains are captured from B-cell repertoires by PCR, and are ligated into phagemids designed for display of either single-chain antibodies (scFv) or Fabs [180]. The main drawback to such libraries, as compared to standard B-cell immortalization techniques, is that original light chain/heavy chain pairs are unmatched. Promiscuous chains that fold well with multiple partners tend to predominate [181, 182]. This has led to efforts to create combinatorial libraries by chain shuffling [183, 184], and to capture the original pairing by single cell PCR [185]. An alternative to native antibody libraries uses randomized CDRs. This randomization can undirected or directed.

Undirected randomization. The first step in affinity maturation of a protein or peptide with known binding characteristics is diversification. There are two basic strategies to randomize libraries in an undirected fashion: low fidelity PCR, and passage in a mutagenic host. Low-fidelity PCR has the advantage of introducing errors only in the portion being amplified. Passage in a mutator strain, such as XL-1 Red [186-188], will introduce mutations throughout the plasmid, potentially introducing changes in promoter regions that might give a growth advantage to library members, or an avidity advantage when expression is increased, thereby increasing valency of display. As mentioned previously, with mammalian display there is also an analogous option of using a cell line with AID. When the target for hypermutation is an antibody, the system for somatic hypermutation in stimulated B cells has the advantage of biasing its mutagenesis towards sequence found in variable chain regions WRCH (W = A/T, R = A/G, H = A/C/T) [189, 190]. Codon-usage biasing toward these sequences could potentially be used to promote mutagenic bias in presentation of other proteins and peptides.

Directed randomization. Mutations at specific locations in a POI can be directed using PCR primers with mixed bases. Most frequently this has been done using arbitrary equimolar mixes of bases. These are also referred to as "machine" mixes, as mixing is performed at time of synthesis by simultaneous injection of the desired bases. Such "hard-randomization" techniques are limited because to have full representation in a a library, the number of positions that can be subject to randomization is small (5 or 6 for phage display). This is generally only useful if one has a very good notion of which positions are likely to produce productive changes. An alternative method is "soft-

randomization", which permits lower frequency changes over a greater number of residues.

In soft-randomization, diversity is introduced by constructing primers with positions biased towards a template or starting-point sequence, mixed with a defined percentage of the other three bases in the mix [191-195]. This is also referred to as base "doping". For example, four mixes may be prepared each containing 85% of one base and 5% each of the other three bases (see Chapter 3). These mixes are used to encode soft-randomized positions, providing a bias towards a starting point or template amino acid. Most mutations introduce change of a single position in a codon. The genetic code is organized such that single point mutations are more likely to be conservative. For example, most codons with T in the first position are hydrophobic, all aromatics but histidine have T in the second position, etc. As a consequence, it is possible to distribute modest changes along the length of the POI without introducing a large number of deleterious mutations. Such a method is particularly useful in an iterative maturation process, where the template for each round of randomization can be based on the best output of the previous round (see Chapter 3). Finally, it is also possible to use both techniques simultaneously, as seen in Dennis et al. and Bahudhanapati et al. [192, 194].

Summary. The battle between HIV-1 and its host plays out most visibly in the constant evolution of Env in competition with the maturation of the antibody repertoire. Surface-display technologies allow us to recapitulate some features of this process. In the following sections we will introduce a new phagemid system combining some of the

advantages of prokaryotic and mammalian systems. In Chapter 3 we will demonstrate the utility of this system by improving a CD4-mimetic peptide. In Chapter 4 we describe the use of this peptide as a fusion with a CCR5 mimetic to create a double-mimetic peptide. Potential future improvement to these systems will be discussed in a final chapter.

CHAPTER 2

Direct expression and validation of phage-selected peptide variants in mammalian cells

Acknowledgements:

The material in this chapter is derived from the published work:

Quinlan BD, Gardner MR, Joshi VR, Chiang JJ, Farzan M. Direct Expression and Validation of Phage-selected Peptide Variants in Mammalian Cells, J. Biol. Chem. 2013 Jun;288:18803-18810

BDQ conceived and executed the creation of this vector.

To whom correspondence should be addressed: Brian D. Quinlan, Department of Infectious Diseases, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL, USA. Tel: (561) 228-2300. Fax: (561) 228-2299. E-mail: bquinlan@scripps.edu, and to Michael Farzan, Department of Infectious Diseases, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL, USA. Tel: (561) 228-2300. Fax: (561) 228-2299. E-mail: mfarzan@scripps.edu.

2. A. ABSTRACT

Phage display is a key technology for the identification and maturation of high affinity peptides, antibodies, and other proteins. However limitations of bacterial expression restrict the range and sensitivity of assays that can be used to evaluate phage-selected variants. To address this problem, selected genes are typically transferred to mammalian expression vectors, a major rate-limiting step in the iterative improvement of peptides and proteins. Here we describe a system which combines phage-display and efficient mammalian expression in a single vector, pDQ1. This system permits immediate expression of phage-selected genes as IgG1-Fc fusions in mammalian cells, facilitating the rapid, sensitive characterization of a large number of library outputs for their biochemical and functional properties.

2. B. INTRODUCTION

Phage-display technology is frequently used to identify protein variants that will ultimately be produced in mammalian cells or expressed *in vivo* [193, 196, 197]. However, bacterially expressed proteins selected as fusions with a phage coat protein do not always express or retain their function in mammalian cells, and in standard phage-display protocols, these defective proteins are retained throughout the selection process [196, 198]. Moreover, peptides too small to be expressed by themselves are typically first evaluated on the surface of the phage by ELISA, but this approach is quantitatively imprecise and retains artifacts from the original selection [191, 199]. This has led to the exploration of alternative display methods, such as yeast or mammalian cell-surface

display [169, 170, 200]. However, the library sizes possible with these approaches, and thus the complexity of the sequence space that can be probed, are orders of magnitude lower than that routinely achieved with phage libraries. To circumvent these difficulties while retaining the power of the phage display method, library outputs can be subcloned to generate fusion proteins, a time-consuming step that limits the number of outputs that can be so evaluated [193, 196, 197, 201, 202]. Expression in bacteria also precludes use of certain fusion proteins, notably those with antibody Fc domains. Fc domains facilitate the use of a broad set of commercial tools for purification, immunoprecipitation, flow cytometry, and functional studies. Ideally, one would incorporate such studies early in the validation of phage-library outputs [203]. Accordingly, we developed a vector that expresses library variants as phage pIII coat-protein fusions in bacterial cells and as fusions with the human IgG1 Fc domain in mammalian cells. This was achieved by inserting the machinery of bacterial expression and phage display within the introns of a mammalian expression vector.

2. C. METHODS

pDQ1 vector construction. pDQ1, represented in Fig. 2.1A, was constructed by inserting an Fc-fusion expression cassette into a pUC-derived plasmid. A segment following the CMV promoter was replaced with sequence encoding the IGH1 signal peptide. The 5' end of the signal peptide intron was then replaced with sequence encoding the Lac operon promoter, a ribosome-binding site and the STII* signal peptide, followed by a CD4-mimetic peptide. The nucleotide and amino acid sequence of the

STII* signal peptide is detailed in Fig. 2.1. Finally, a segment from a pComb3 vector [180] encoding the HA tag and the pIII fusion was inserted into the IgG1 hinge-CH1 intron immediately following the VH1 splice donor.

Use of the vector. The current version of the vector has been used for display and expression of peptide or scFv, which expresses as scFv-Fc in mammalian cells. Peptide sequence is introduced using inverse PCR. A minor modification of the vector placed an scFv of the CD4-binding site antibody b12 in the vector (not shown) with the restriction sites SpeI-BamHI for exchange of scFvs. A full implementation of the pDQ1 vector is described in Chapter 3.

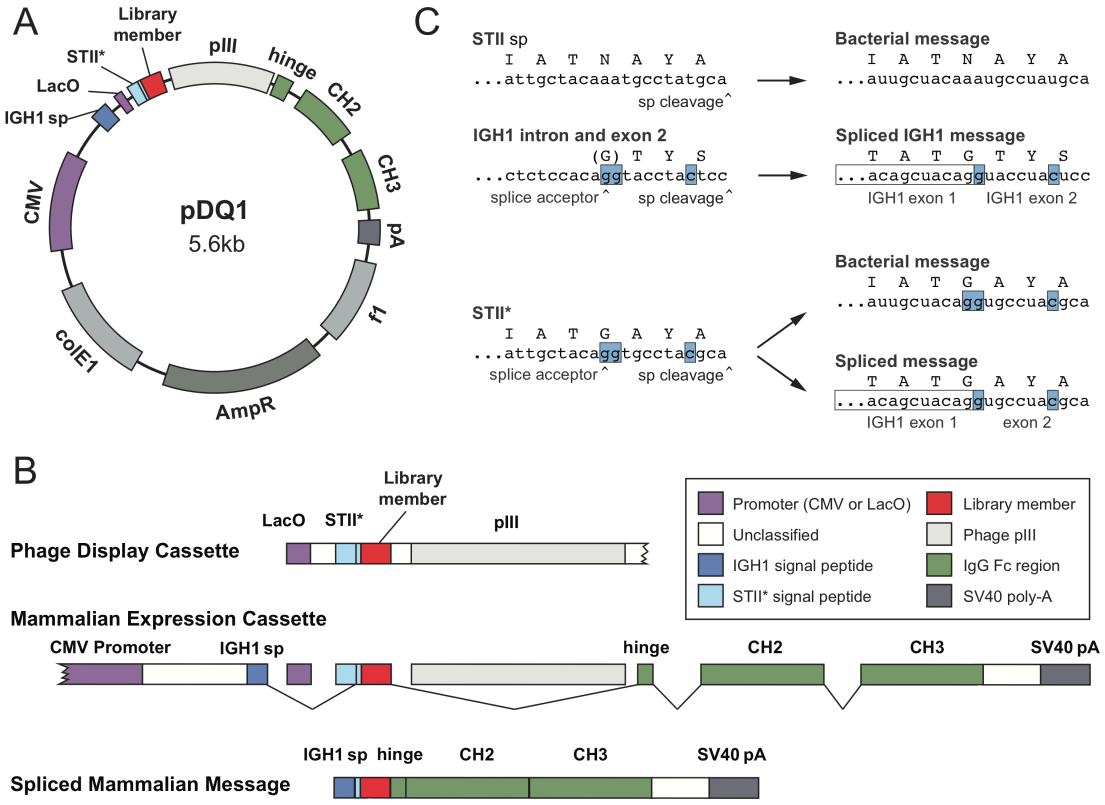

Figure 2.1. The design of pDQ1, a vector for phage-display and mammalian-cell expression of protein-Fc fusions. **A,** A map of pDQ1 is shown, with mammalian (CMV) and bacterial (LacO) promoters indicated (purple), as are exon 1 of the heavy chain of IgG1 (IGH1) encoding most of the IgG1 signal peptide (blue), STII*, encoding a form of the bacterial signal peptide STII (heat-stable enterotoxin II) modified to include a splice acceptor (light blue), the diversified library member (red), the phage pIII protein, individual exons comprising the Fc domain of human IgG1 (green), and an SV40 polyadenylation termination signal (gray). **B,** Bacterial and mammalian expression cassettes. Phage-display uses a cassette including the LacO promoter, and sequence encoding the STII* signal peptide and a library component fused to the phage pIII protein. In contrast, mammalian expression utilizes a CMV promoter, and encodes a chimera of the IgG1 signal peptide and STII*, the same library component, and the Fc domain of IgG1, terminated with an SV40 polyadenylation signal. Note that in mammalian cells splicing excises the LacO promoter, most of STII*, and the pIII gene. **C,** Modifications of the bacterial STII signal peptide necessary for expression in mammalian cells. The DNA and amino-acid sequences of the C-terminus of the STII signal peptide are shown at top, with the signal peptidase cleavage site indicated. This signal peptide functions in bacteria but not in mammalian cells. Below that, intronic sequence and the 5' region of IGH1 exon 2 are shown. This region encodes the C-terminus of the IgG1 signal peptide. The 3' region of the splice acceptor and the site of signal peptidase cleavage are indicated. In mammalian cells, IGH exon 2 is spliced to exon 1, encoding the complete IGH1 signal peptide. Underneath, sequence encoding the C-terminus of the STII* signal peptide is shown. Blue highlighting indicates nucleotides from exon 2 of IGH1 that introduce an efficient splice acceptor while retaining signal peptide function in both bacterial and mammalian cells. In bacteria, the entire STII* signal peptide is used, whereas in mammalian cells, exon 1 of IGH1 is spliced to a short 3' region of STII*.

2. D. RESULTS

Design of a dual expression vector. The vector, shown in Fig. 2.1A, exploits the fact that splicing occurs only in eukaryotic cells. Splice donors and acceptors were engineered to excise domains specific to phage display and thereby encode a protein-of-interest fusion with the Fc domain of human IgG1 (Fig. 2.1B). A first splice links regions encoding a mammalian signal peptide to one encoding the protein under selection, excising sequence encoding a bacterial promoter and signal peptide. A second splice links the gene for the protein under selection to sequence encoding the hinge region of IgG1, excising the gene for phage pIII. Thus in bacteria the vector functions like standard phagemid vectors (e.g. pCOMB3) used for phage display [180], whereas when transfected into mammalian cells, protein-Fc fusions are expressed. Our design required that the bacterial and mammalian signal sequences share four C-terminal amino acids and encode an efficient splice acceptor. No native bacterial or mammalian sequence met these requirements. However, using splice site and signal peptide prediction programs [204, 205], we combined elements from the bacterial STII signal peptide and the mammalian IgG1 signal peptide to generate a novel signal peptide (STII*) that allowed for efficient expression in both systems (Fig. 2.1C). The resulting vector, pDQ1, expressed Fc-fusion products in mammalian cells as efficiently as the original pCDM8-derived vector from which the Fc gene was cloned (not shown) [206], and was used to produce all the Fc-fusion proteins described here and in Chapter 3.

2. E. DISCUSSION

Phage display remains a powerful means of improving bioactive peptides and proteins that in many cases will ultimately be produced in mammalian cells. Current approaches for validating library outputs limit the number of outputs that can be reasonably evaluated or the precision and range of assays that can be applied. The pDQ1 vector described here permits early selection for efficient mammalian expression and expands the range of available validation techniques. It thus facilitates immediate, sensitive characterization of a large number of library outputs for their biochemical and functional properties. One potentially useful adaptation of this system will be the introduction of an Fab library into the vector, facilitating direct production and characterization of full-length antibodies in mammalian cells. With minor modifications, this system might also be adapted for mammalian cell surface-display, allowing for a preliminary selection in phage followed by subsequent rounds of selection by cell-surface display on subsets of manageable size. This system can also be used without phage display, for example when determining a production technology for a particular biologic or when the role of a eukaryotic post-translational modification is investigated. However, it is especially useful for phage-display applications. Because both phage-display and mammalian expression from pDQ1 remain as efficient as with single-function vectors, we anticipate that mammalian expression will become a standard capability of vectors used for phage-based improvement of peptides and proteins.

CHAPTER 3

Maturation of a CD4-mimetic peptide

Acknowledgements:

The material in this chapter is derived from the published work:

Quinlan BD, Gardner MR, Joshi VR, Chiang JJ, Farzan M. Direct Expression and Validation of Phage-selected Peptide Variants in Mammalian Cells, J. Biol. Chem. 2013 Jun;288:18803-18810

Experiments designed by MF and BQ. Receptor complementation assay by MG (Fig 5B). Neutralizations with CD4mim6W-Ig performed by VJ (Fig 5C).

To whom correspondence should be addressed: Brian D. Quinlan, Department of Infectious Diseases, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL, USA. Tel: (561) 228-2300. Fax: (561) 228-2299. E-mail: bquinlan@scripps.edu, and to Michael Farzan, Department of Infectious Diseases, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL, USA. Tel: (561) 228-2300. Fax: (561) 228-2299. E-mail: mfarzan@scripps.edu.

3. A. ABSTRACT

The pDQ1 is a newly described vector integrating phage display and mammalian expression. We demonstrate the utility of this system by improving the ability of a CD4-mimetic peptide to bind the HIV-1 envelope glycoprotein and neutralize HIV-1 entry. We further improved the potency of the resulting peptide, CD4mim6, by limiting its ability to induce the CD4-bound conformation of the envelope glycoprotein. Thus CD4mim6 and its variants can be used to investigate the properties of the HIV-1 envelope glycoprotein and pDQ1 can accelerate the discovery of new peptides and proteins through phage display.

3. B. METHODS

Library design and assembly. Primers were designed using four hand-mixes, each containing 85% of the template nucleotide and 5% of the remaining three nucleotides. Library 1 was constructed using CD4mim2 as a template. In all cases positions 6, 10, 21-24 and 26 were held fixed as indicated in Fig. 3.1B. The remaining positions were soft randomized with the appropriate hand mix for the first two codon positions, and 50:50 G:C mix was used for the wobble position. Library 2 was constructed using CD4mim4 as a template. In this iteration, the wobble position was soft-randomized for amino-acids encoded by fewer than four codons (K, N, H, Q, D, E, M, I, Y and W) and a 50:50 G:C mix was used in the remaining cases (see Fig. 3.2). The 3' end of the primer extended

outside of the randomization region, and was of sufficient length to have a melting temperature ≥ 72°C. Primers were phosphorylated with T4 PNK (New England Biosciences) and used for inverse PCR in 96 well plates with Phusion Flash (Finnzymes). The PCR product was pooled, column purified (Qiagen) and ligated with T4 DNA Ligase (New England Biolabs). The ligation product was precipitated using yeast tRNA carrier [207] and transformed into MC1061 F' (Lucigen), titered on LB agar Ampicillin, and spread onto multiple 140 mm LB Agar Carbenicillin/Tetracycline/glucose plates.

Phage production and panning. Libraries were resuspended from bacterial plates using Super Broth + 0.5% glucose and transduced with VCSM13 helper phage (Stratagene) as describe in Rader et al. [208] Briefly, expression was induced overnight with SB autoinduction media (yeast extract 2%, tryptone 3%, MOPS 1%, glycerol 0.5%, $MgCl_2$ 1 mM, glucose 0.05%, lactose 0.05%) containing Carbenicillin (100 mg/L) and Kanamycin (50 mg/L). The supernatant was cleared by centrifugation, passed through a 0.45 μm filter and precipitated with PEG/NaCl + EDTA. The pellet was resuspended in cold Tris-buffered saline (TBS) and EDTA was added. gp120-Ig or Fc-only protein A or protein G magnetic beads (Invitrogen) were prepared by incubation of gp120-Ig or Fc control with beads for 1 hr at 4°C in TBS-TC (0.01% Tween 20, 1% sodium caseinate) followed by two washes. Freshly prepared phage were incubated 10 min with beads and washed 10-20 times with TBS-T. Retained phage were eluted with freshly prepared 1% Trypsin in TBS, and eluate was used to transform cultures of NEB $F'I^q$, and dilutions of gp120 and Fc-control cultures were plated to monitor enrichment. After 1 hour of recovery in Super Broth + glucose, cultures were plated as before. Colonies from the titer plates were sequenced to monitor library diversity, and the expansion and panning

process was repeated for a total of 3 or 4 pans until the library size was reduced to approximately 30 unique clones out of 200 sequences.

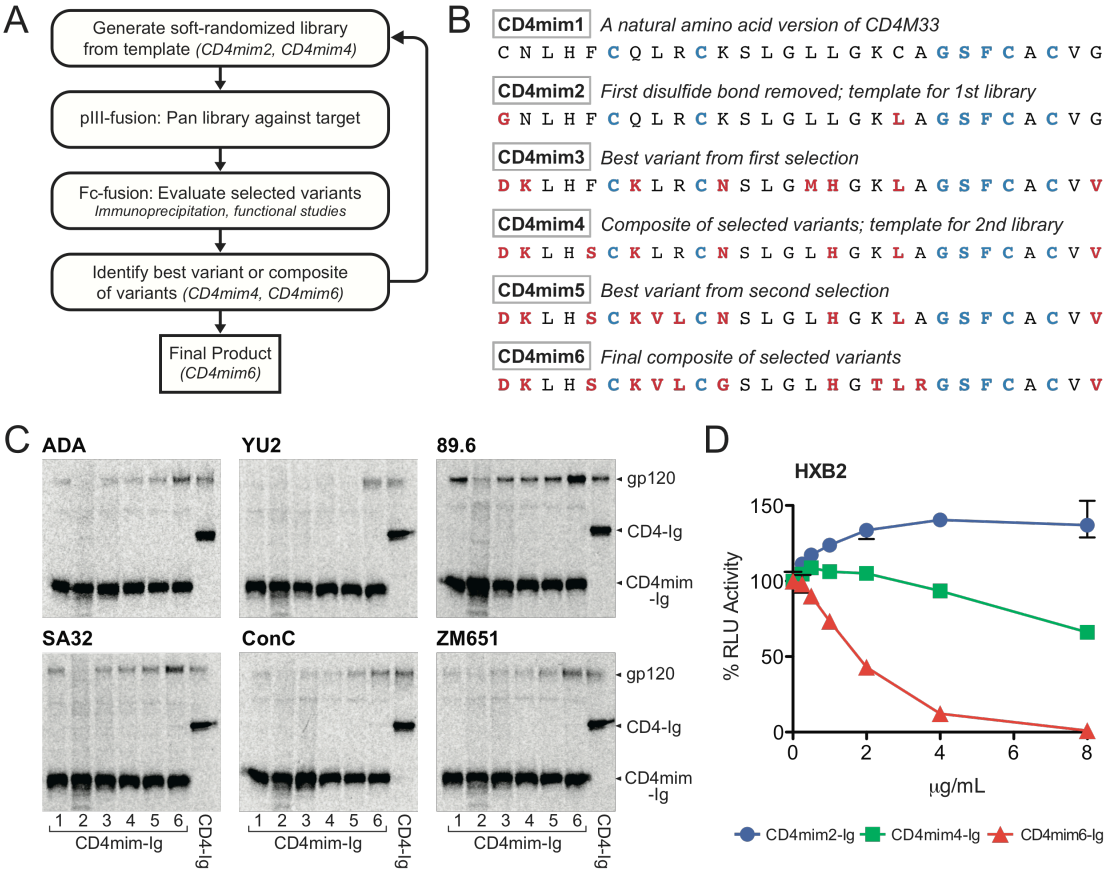

Figure 3.1. Incorporation of mammalian cell expression and functional assays into the workflow of phage display. A, The streamlined phage display workflow used in this study. Plasmids selected in panning against target were isolated and used to transfect 293T cells. Library outputs were produced as Fc-fusion proteins, which facilitated immunoprecipitation, flow cytometry, and neutralization studies. Elements of high performing variants were combined and reassayed. A best composite variant was used as a template for a subsequent round of phage-selection and characterization of Fc fusions. **B,** The sequences of the CD4mim variants are shown with description. Blue indicates seven residues held fixed through the process, red indicates changes from CD4mim1, a natural amino-acid form of a previously described CD4-mimetic peptide. **C,** An example immunoprecipitation study used to validate library outputs. Here, isotopically labeled CD4mim1-6 and CD4-Ig are compared for their abilities to precipitate labeled gp120 of the indicated isolate. ADA, YU2, and 89.6 are clade B isolates; SA32, ConC, and ZM651 are clade C isolates. Experiment is representative of three with similar results. **D,** An example of a functional study used to characterize library outputs. Here, CD4mim2-Ig, CD4mim4, and CD4mim6-Ig are compared using a TZM-bl HIV-1 neutralization assay. Error bars represent a range of triplicates.

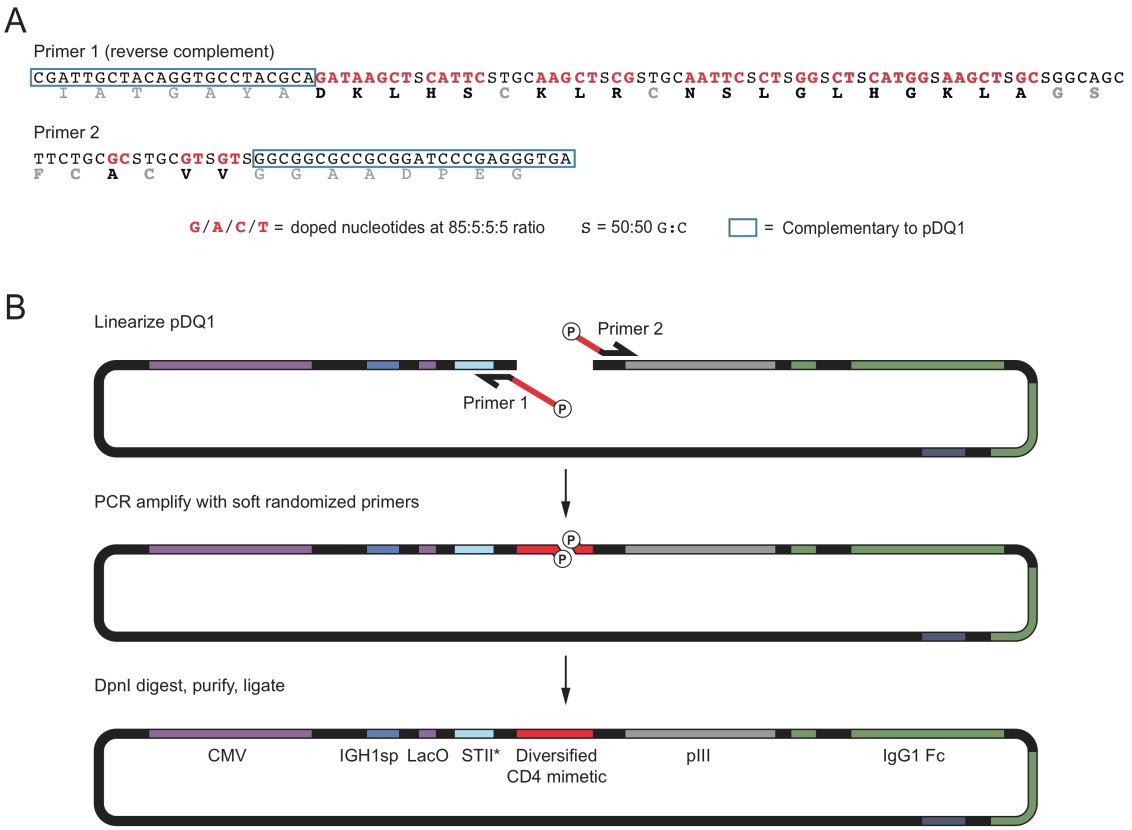

Figure 3.2. The generation of a soft-randomized library using pDQ1. A, Primers used to generate a soft-randomized library of CD4mim variants based on a template of CD4mim4. Regions used for priming and complementary to the STII* signal peptide (Primer 1) or the second splice donor (Primer 2) are boxed. Bold red nucleotides indicate hand-mixes at an 85:5:5:5 ratio favoring the original template nucleotide. S indicates a 50:50 ratio of G:C, used for wobble positions when an amino acid is encoded by four or more distinct codons. Amino acids in bold black are soft-randomized, whereas gray amino acids are held fixed. **B,** Library synthesis. The primers shown in A were phosphorylated and used to amplify a linearized pDQ1 vector. The 5.6 kB PCR product was digested with Dpn I to eliminate the original vector, and the full-length product was isolated and ligated. Colors correspond to elements of text Fig. 2.1.

Immunoprecipitation. Miniprep DNA prepared from sequenced colonies was transfected into 12-well plates of 50% confluent 293T cells at 1 μg DNA/well by standard calcium phosphate technique. HIV-1 gp120 constructs were transfected into 6 well plates. 12 hours post-transfection, media was replaced with cysteine/methionine-negative DMEM containing ExpressLabel (Cole Parmer) and 5% dialyzed calf serum

(Invitrogen). Supernatants were collected 24 hours later, and were cleared by centrifugation 2 min at 3,000 rpm, followed by 1 min at 10,000 rpm and complete protease inhibitor cocktail (Roche) was added. Immunoprecipitation assays were conducted by incubating normalized quantities of Ig fusions and gp120 supernatants with protein A beads for 2 hrs at 4°C with rocking. Beads were washed 3 times with PBS-Tween (0.01%) and protein was eluted by heating 10 min at 70°C in NuPage sample buffer. Samples were run on NuPage bis-tris gels, treated with methanol/acetic acid fixative and dried on Whatman paper. Radiolabeled protein was quantified with Phosphoimager (Fuji).

Protein production. Maxiprep DNA was prepared from 250 mL overnight LB cultures of DH5α carrying the pDQ1-CD4mimetic plasmid using PureLink maxiprep kits (Invitrogen) according to the manufacturer's instructions. 293T cells in 140 mm plates were transfected with 25 μg/plate at 50% confluency by standard Ca Phosphate transfection. At 12 hrs post-transfection, 10% FCS-DMEM media was replaced with serum free 293 Freestyle media (Invitrogen). Media was collected 48 hrs later, debris was cleared by centrifugation for 10 min at 1,500g and was filtered in 0.45 μm filter flasks (Millipore). Complete protease inhibitor cocktail (Roche) was added to the filtered supernatants. 500 μl bed volume of Protein A sepharose beads (GE Healthcare) were added and were agitated 4°C overnight. The bead/media mixture was collected by gravity flow column (Biorad) and was washed with 30 mL PBS (Lonza) + 0.5M NaCl (0.65M NaCl final) followed by 10 mL PBS. Protein was eluted with 5 mL 2M arginine pH4 into 1 mL 1M Tris pH 7.5. Buffer was exchanged for PBS and protein was concentrated to 1 mg/ml by ultrafiltration (Amicon Ultra) at 3,000g.

Surface plasmon resonance studies. Surface plasmon resonance biosensor data were collected on a Biacore 3000 optical biosensor (GE Healthcare). gp120 of the clade C Du151 isolate (Immune Technology Corp.) was coupled to cell 2 or cell 4 of a CM5 chip using the Amine Coupling Kit (Biacore). Kinetic data was collected at several concentrations of mimetic-Fc fusion proteins, with two repeated concentrations. Surface regeneration was by injection of 1M potassium thiocyanate, 1% CHAPSO. Data were analyzed using BIAevaluation software (GE Healthcare) and fit to a 1:1 Langmuir binding model.

HIV-1 neutralization studies. Pseudotyped virus was produced by coexpression of envelope glycoproteins of the indicated HIV isolates, or the VSV-G protein, with NL43ΔEnv. 293T cells at 50% confluency in T175 (Falcon) flasks were transfected with 25 μg of plasmid encoding envelope glycoprotein and 45 μg of NLΔEnv by the standard Calcium Phosphate technique. 10% FCS DMEM media was changed at 12 hrs and media was collected at 48 hrs. Viral supernatants were cleared by centrifugation for 10 min at 1,500g, passed through a 0.45 μm syringe filter (Millipore), and stored at -80°C. Neutralization by Fc-fusion constructs was performed according to the protocol described in Li et al. [209] In brief, varying concentrations of CD4-mimetic-Ig or free CD4-mimetic peptide (NeoBioLab) were incubated with virus at a volume of 100 μl per well in 96 well plates for one hour at after which 10,000 cells/well of TZM-bl cells in 100 μl of media were added. Plates were incubated 72 hours at 37°C after which 100 μl of media was replaced with freshly prepared Brite-Lite reagent (Perkin-Elmer), and luminescence data was collected on a Victor3V (Perkin Elmer). For peptide neutralizations, 10,000 cells per well were first plated in 96 well plates. 12 hours later, peptide and pseudoviruses

were mixed in V-bottom 96-well plates, incubated one hour and then added to the adhered cells. The media was changed after 2 h incubation at 37°C. All neutralizations were performed in triplicate.

Infection of CCR5-positive, CD4-negative cells. CF2th-CCR5 cells were incubated with pseudoviruses generated with pNL4-3.Luc.R-.E-bearing the indicated HIV-1 envelope glycoprotein and encoding firefly luciferase in the presence of increasing amounts of peptide forms of CD4mim6, CD4mim6W, or soluble CD4. Infection was measured as in neutralization studies.

Staining of cell expressed HIV-1 envelope glycoprotein. Cells expressing HIV-1 envelope glycoprotein trimers were prepared by transient transfection of plasmids expressing envelope glycoproteins truncated in the C-terminal cytoplasmic domains to facilitate surface expression (gp160-ΔCT). T75 flasks of 293T at 50% confluency were transfected with 20 μg of plasmid encoding gp160-ΔCT and 3 μg of plasmid encoding tat by the calcium phosphate technique and incubated overnight. DMEM + 10% FCS was changed at 18 hours post transfection. Cells were collected 36 hours post transfection by treatment with non-enzymatic cell dissociation solution (Sigma), followed by centrifugation 5 minutes at 1200 rpm. Cells were resuspended in flow cytometry buffer (PBS + 2% Donor Goat Serum + 0.1% sodium azide). Cells were incubated with Ig-fusion proteins or the HIV-1 neutralizing antibodies E51 or 2G12 diluted to varying concentrations in FACS buffer for 45 min on ice, followed by two FACS buffer washes. APC-conjugated anti-human Fc-gamma (Jackson ImmunoResearch) diluted in FACS buffer was added and cells were incubated on ice for 30 minutes. Cells were then washed

2 times with flow cytometry buffer and PBS then fixed for flow cytometric analysis with 1% paraformaldehyde in PBS. Flow cytometery data were analyzed using FlowJo.

3. C. RESULTS

Construction of a 'soft-randomized' library. In a standard phage-display workflow, bacteria culture supernatants, periplasmic extracts, or precipitated phage are used to initially validate and compare selected phagemid clones. The alternative workflow used here integrates mammalian expression of selected variants, as represented in Fig. 3.1A. As indicated, peptide Fc-fusion proteins were produced directly from pDQ1 phagemid clones in mammalian cells, and these Fc-fusion proteins were assayed with a range of techniques.

Here, for example, immunoprecipitations of HIV-1 gp120, the target protein, were performed using metabolically labeled supernatants of transfected 293T cells. In parallel, promising peptide-Fc variants were compared using flow cytometry, HIV-1 neutralization studies, and surface plasmon resonance. Elements from optimal variants were then combined, and the composite peptide that best bound gp120 and neutralized HIV-1 then served as the template for a new phagemid library and a subsequent round of selection.

We began with a previously described CD4-mimetic peptide that included three unnatural amino acids [138, 146, 210]. These amino acids were reverted to their closest natural form, introducing a cysteine at position 1, a phenylalanine at position 23, and a

valine at position 27 to generate CD4mim1 (Fig. 3.1B). In order to facilitate N-terminal fusions of this peptide with CCR5-mimetic peptides [78, 112], and to expand the conformational space available to peptide variants, the first of three disulfide bonds was eliminated (CD4mim2). Note that CD4mim2 bound gp120 with lower affinity than CD4mim1, presumably due to the loss of this disulfide bond. CD4mim2 was used as a template to generate a first library of CD4mim variants. CD4mim4, the final composite product of our first selection cycle, was used as a template for a second library and another round of selection. Both libraries were 'soft-randomized' from their respective templates.

Soft randomization is a diversification technique that permits the introduction of a small number of amino-acid changes (e.g. 5) distributed throughout a larger peptide (e.g. the 28 amino-acids of CD4mim2 and CD4mim4). This approach allows modest changes along the length of the peptide without introducing a large number of mutations likely to be deleterious [191, 193, 195]. Fig. 3.2 details our procedure for library construction using pDQ1 and soft-randomized oligonucleotides. Seven amino acids were held fixed: the four remaining cysteines and a "GSF" region identical to the key gp120-binding residues of CD4 (residues 41-43 [210]), whereas the remaining twenty-one residues were encoded by doped oligonucleotides favoring template amino acids at an 85:5:5:5 ratio. This library was panned against HIV-1 gp120 (those of the clade B isolate ADA and the clade C isolate ZM651, see Methods). Selected CD4mim variants were expressed as peptides-Fc fusions in 293T cells and used to compare output variants for functional activity in immunoprecipitation and HIV-1 neutralization assays (Figs. 3.1C and 3.1D for example). Elements of selected variants were combined and resulting proteins were

again compared. In this way, an optimal peptide was selected as a template for a second soft-randomized library. The procedure was repeated, and an improved peptide, CD4mim6, was selected. Fig. 3.3 details the process by which CD4mim6 was developed as a composite of several library outputs.

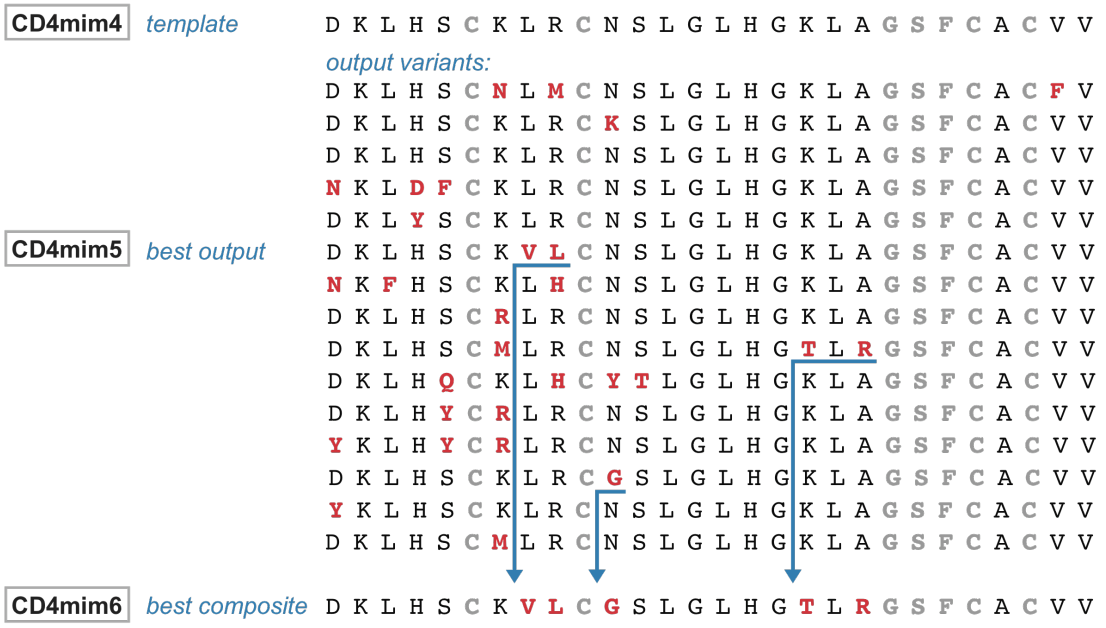

Figure 3.3. The development of CD4mim6. The sequence of CD4mim4, the template for a second soft-randomized library, is shown along with a representative panel of outputs from this library panned against a clade B (ADA) and clade C (ZM651) isolate. Outputs by immunoprecipitation (for example Fig. 2C), flow cytometry (e.g., Fig. 3A) and neutralization studies (e.g. Figs. 2D and 3C) identified best performers whose elements were recombined. The resulting composite variants were reassayed, and an optimal variant, in this case, CD4mim6 was identified. As shown CD4mim6 combines elements of CD4mim5, the best output from this screen, and two additional good outputs.

Enhanced neutralization of HIV-1 by CD4mim6. Because CD4mim1, with three disulfide bonds, outperformed in immunoprecipitation and neutralization assays our starting point, CD4mim2, with two disulfide bonds (Fig. 3.1D and not shown), we compared the properties of our final product, CD4mim6, with CD4mim1, using Fc fusions of both variants. Both peptides bound measurably to 293T-cell surface expressed

HIV-1 envelope glycoprotein trimers (Fig. 3.4A). However, at all concentrations, and for all clade B and C envelopes assayed, CD4mim6-Ig bound more efficiently than CD4mim1-Ig, and similarly to CD4-Ig. Surface plasmon resonance studies with immobilized gp120 (Du151, clade C) made clear that the basis for the efficient binding was due to a slower off-rate, five times slower for CD4mim6-Ig. In fact, CD4mim6-Ig had a modestly slower on-rate, presumably due to its greater flexibility (Fig. 3.4B). More total CD4mim6-Ig bound to immobilized gp120, perhaps indicating that the more flexible CD4mim6-Ig accessed more gp120 conformations than the fixed CD4mim1-Ig.

CD4mim6-Ig also efficiently neutralized clade B and clade C HIV-1 isolates at concentrations (1-2 μg/ml; 20-40 nM) where no neutralization was observed for CD4mim1-Ig (Fig. 3.4C). When CD4mim1 and CD4mim6 were synthesized as free peptides without an Fc domain, CD4mim6 efficiently neutralized HIV-1, again at concentrations where no neutralization was observed from CD4mim1 (Fig. 3.4D). Figures 3.5A and B shows a model of CD4mim6 bound to HIV-1 gp120, with changes from CD4mim1 highlighted.

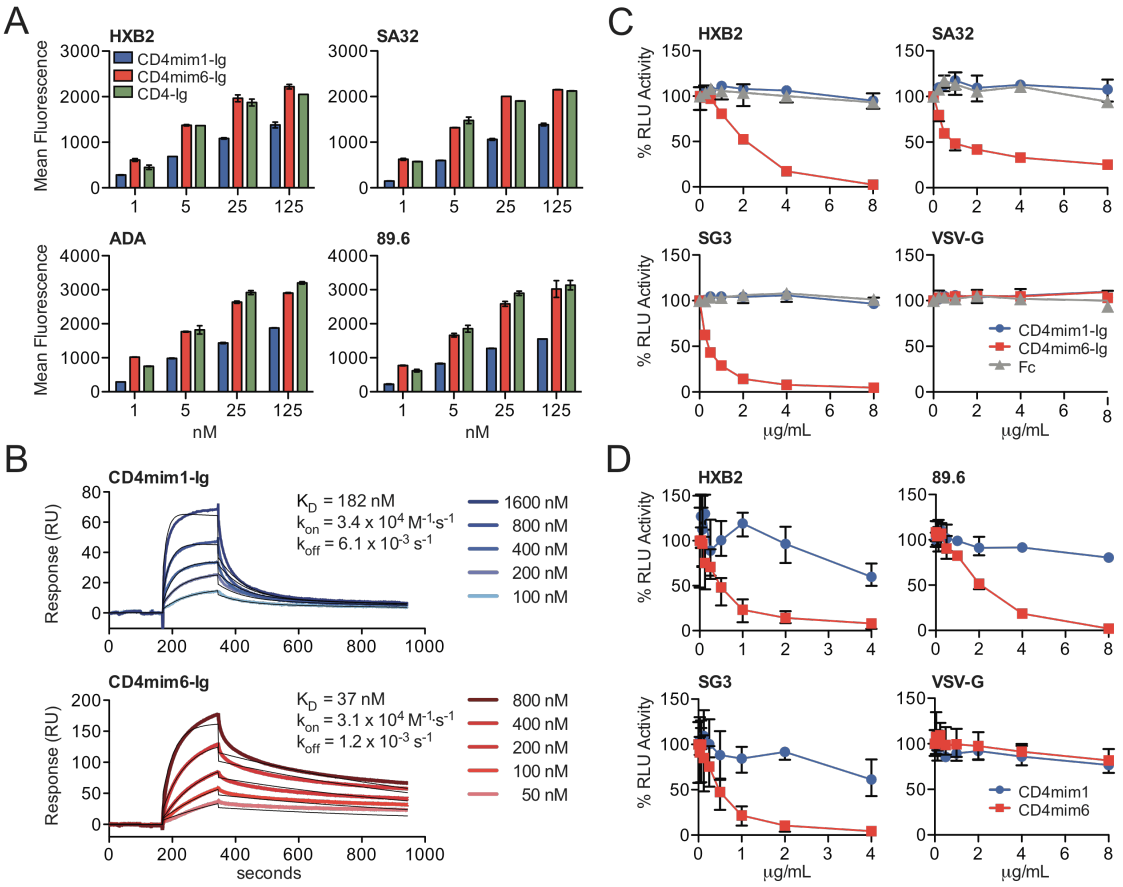

Figure 3.4. Comparisons of CD4mim1-Ig and CD4mim6-Ig. A, Binding of CD4mim-Ig variants and CD4-Ig to 293T cells transfected to express the trimeric envelope glycoproteins of the indicated clade B (HXB2, ADA, 89.6) and clade C (SA32) isolates. Binding was determined by flow cytometry using a goat anti-human secondary antibody. Background binding, determined using Fc domain alone was less than 15 mean fluorescence intensity in each case. Experiment is representative of two with similar results. **B,** Surface-plasmon resonance studies of CD4mim1-Ig and CD4mim6-Ig using immobilized gp120 of the clade C isolate Du151. Results are shown with thick lines. Thin black lines represent global fitting of the data to a 1:1 Langmuir binding model. **C,** CD4mim6-Ig neutralizes clade B (HXB2) and clade C (SG3, SA32) HIV-1 isolates with IC50s of 1-2 µg/ml (20-40 nM), whereas no neutralization was observed for CD4mim1-Ig at 8 µg/ml (160 nM), and HIV-1 pseudotyped with VSV-G protein was unaffected by any Fc fusion. Experiment is representative of three with similar results. **D,** Synthetic monomers of CD4mim1 and CD4mim6 were compared in the same assay. CD4mim6 neutralized the indicated isolates with IC50s in the 0.5-2 µg/ml range, whereas IC50s for CD4mim1 were greater than 4-8 µg/ml. Experiment is representative of two with similar results. Error bars in all cases represent a range of duplicates (A) or triplicates (C and D).

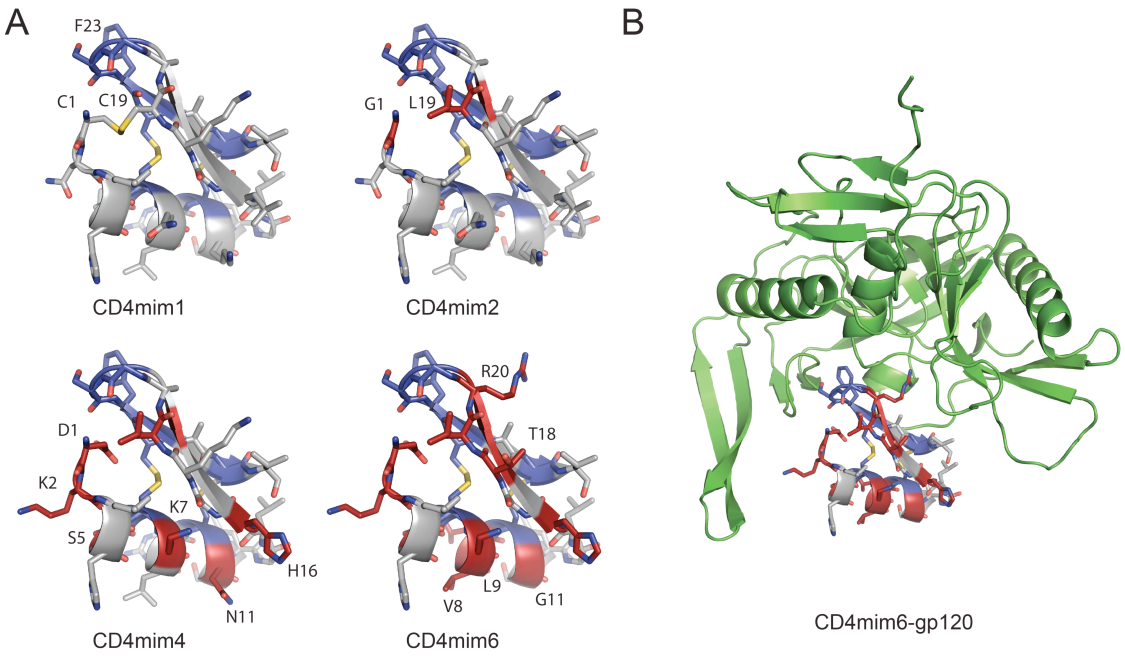

Figure 3.5. A model of CD4mim variants with alterations indicated. A, CD4mim1 is a natural amino- acid form of CD4M33, a CD4-mimetic peptide built from a scorpion toxin scaffold, with three disulfide bonds. Residues held constant during the library construction are indicated in blue. For CD4mim2, cysteines 1 and 19 were altered to glycine and leucine, respectively, shown in red. CD4mim4, the product of the first cycle of phage selection and the template the second phage library has eight total differences from CD4mim1, again indicated in red (see Fig. 2B for changes). CD4mim6, the final product of the second cycle, has twelve such differences. **B,** CD4mim6 modeled into the CD4-binding site of HIV-1 gp120. All models are based on the Protein Data Bank structure 1YYM by Huang et al. (20) and generated with the PyMOL mutagenesis tool.

CD4mim6 as a tool for understanding neutralization of HIV-1. To determine the similarity between CD4mim6 and sCD4, we incubated CD4mim6 with cells expressing the HIV-1 envelope glycoprotein ADA, and measured its ability to promote association with the CD4-inducible antibody E51 (Fig. 3.6A). CD4mim6 increased association of E51 with the envelope glycoprotein, but had no effect on an HIV-1 neutralizing antibody, 2G12, that is unaffected by CD4 association. In parallel, we assayed a CD4mim6 variant,

CD4mim6W, in which phenylalanine 23 was replaced with a tryptophan. Phenylalanine 23 binds a critical pocket of HIV-1 gp120 also occupied by CD4 phenylalanine 43. Occupation of this same gp120 pocket with a tryptophan has been shown to increase the potency of the neutralizing antibody NIH-45-46 [211]. Strikingly, CD4mim6W did not enhance binding of E51 to HIV-1 envelope glycoprotein. We speculated that it would therefore be unable to promote infection when cellular CD4 was limiting, in contrast to both sCD4 and CD4mim6. Indeed with CD4-negative, CCR5-high cells, infection increased with increasing concentration of sCD4 or CD4mim6 (Fig. 3.6B). However, little or no infection was observed at any concentration with CD4mim6W. Consistent with these observations, CD4mim6W more potently neutralized HIV-1 than CD4mim6 (Fig. 3.6C). Thus the potency of CD4mim6 and perhaps CD4-binding site antibodies may be attenuated by their tendency to promote infection when cellular CD4 is limiting.

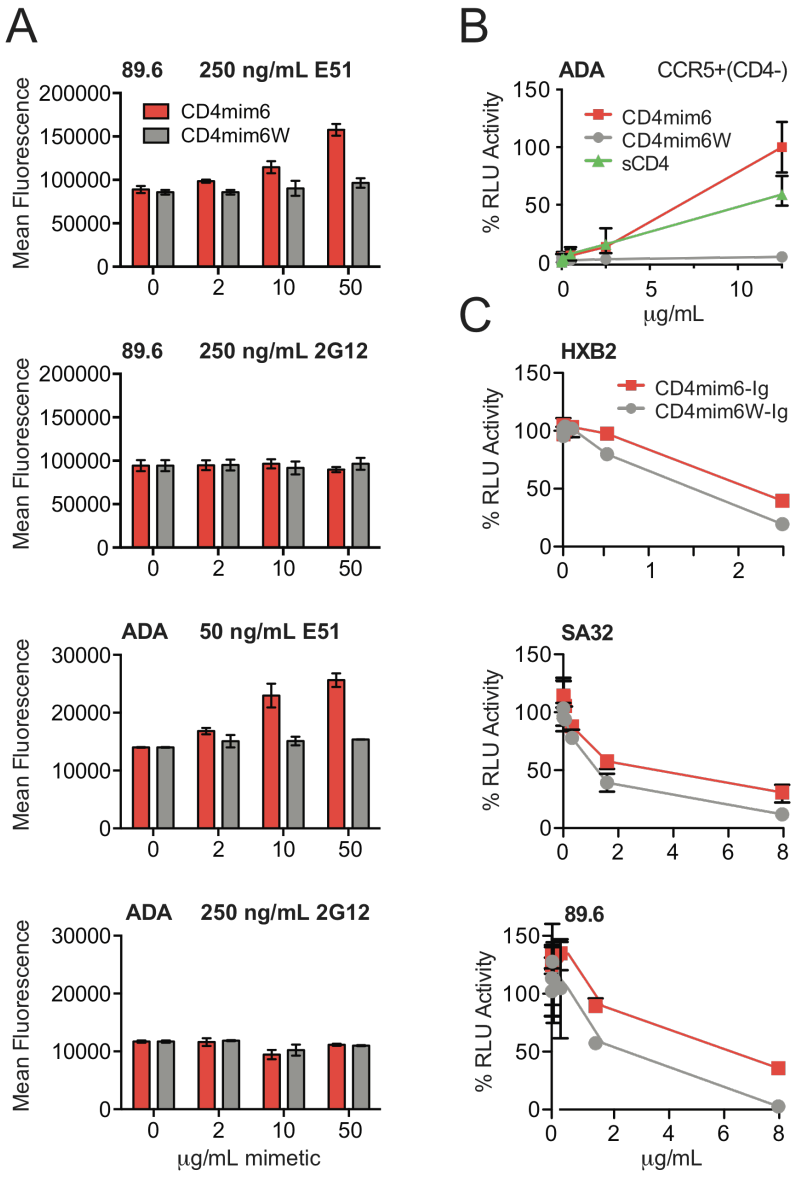

Figure 3.6. CD4mim6, but not CD4mim6W, induces the CD4-bound conformation of the HIV-1 envelope glycoprotein. A, 293T cells transfected to express the envelope glycoprotein of the clade B isolate ADA were incubated with the indicated concentrations of CD4mim6 or CD4mim6W, a variant in which phenylalanine 23 was replaced with a tryptophan. Cells were then incubated with an anti-gp120 antibody, either the CD4-inducible antibody E51(250 ng/mL for 89.6; 50 ng/mL for ADA) or the anti- glycan antibody 2G12 (250 ng/mL). Binding was determined by flow cytometry using a goat anti-human secondary antibody. Background binding, determined using Fc domain alone was less than 15 mean fluorescence intensity in each case. Experiment is representative of two with similar results. **B,** CCR5- high, CD4-negative Cf2Th cells were incubated with an ADA pseudovirus encoding firefly luciferase in the presence of the indicated concentrations of CD4mim6, CD4mim6W, or soluble CD4 (sCD4). Infection, measured as luciferase activity (RLU) was determined as in Fig. 3.4C. **C,** Neutralization studies of CD4mim6-Ig and CD4mim6W-Ig as described in Fig. 3.4C.

3. D. DISCUSSION

The CD4-mimetic peptide generated in these studies is unlikely by itself to serve as a therapeutic. However, a small peptide that mimics CD4 can be especially useful as a tool for understanding neutralization of HIV-1. Its size ensures that it will not sterically interfere with binding of antibodies to the HIV-1 envelope glycoprotein. It therefore can be used with various antibodies to compare the CD4-bound and unliganded states of the envelope glycoprotein. Further study of CD4mim6W may also deepen our understanding of the different conformational states available to gp120 and the envelope glycoprotein trimer.

Although the therapeutic potential of CD4mim6 alone is limited, it may be more potent as a fusion with other inhibitors. CD4mim1 has been fused at its carboxy-terminus with tyrosine-sulfated CCR5-mimetic peptides to make a 'double-mimetic' peptide with considerably higher potency than either peptide alone [78]. CD4mim6 possess two advantages over CD4mim1 in this context. First, it binds gp120 with higher affinity and neutralizes HIV-1 more efficiently than CD4mim1. Second, some steric interference between the amino-terminus of CD4mim1 and gp120 has been observed [138], precluding N-terminal fusions with this mimetic. By eliminating the first disulfide bond of CD4mim1, the amino-terminus of CD4mim6 can now be readily fused to a CCR5-mimetic peptide, positioning the latter mimetic much closer to its binding pocket. These kinds of constructs induce the CD4-bound conformation of the trimer, and therefore can potentially synergize with gp41-derived peptides such as T20/enfuvirtide. Further improvement of both double and single mimetic peptides, and of neutralizing antibodies that target HIV-1 and other pathogens, can be significantly accelerated using

the pDQ1 system described here.

CHAPTER 4

A double-mimetic peptide efficiently neutralizes HIV-1 by bridging the CD4- and coreceptor-binding sites of gp120

Acknowledgements:

The material in this section is derived from work in submission:

Quinlan BD, Joshi VR, Gardner MR, Farzan M; A double-mimetic peptide efficiently neutralizes HIV-1 by bridging the CD4- and coreceptor-binding sites of gp120; 2013

Experiments designed by MF and BQ. Neutralizations for Fig 4.2B and Fig 4.4B by VJ. Neutralizations for Fig 4.2A by BQ and VJ.

Corresponding author. Mailing address: Michael Farzan, The Scripps Research Institute-Scripps Florida, 130 Scripps Way, Jupiter, Florida, 33458. Phone: 561-228-2300. E-mail: mfarzan@scripps.edu

4. A. ABSTRACT

We show that fusions of CCR5- and CD4-mimetic peptides, expressed as immunoadhesins, neutralize HIV-1 more efficiently than CD4-Ig or mixtures of immunoadhesin forms of each peptide. Specifically, double-mimetic peptides with linkers of 11 amino acids in which the CCR5-mimetic component precedes the CD4-mimetic component were more efficient than constructs with shorter linkers or a reverse orientation. These fusion proteins bridge the CD4- and CCR5-binding sites of HIV-1 gp120 to achieve their greater potency.

4. B. INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) entry requires cellular expression of CD4 and a coreceptor, principally CCR5 or CXCR4 [10, 11, 13, 212]. Small CD4-mimetic peptides, based on a scorpion-toxin scaffold, bind the HIV-1 envelope glycoprotein (Env) and inhibit HIV-1 entry [145, 146, 213]. These peptides induce the CD4-bound conformation of gp120 and, like soluble CD4, can promote infection when cellular CD4 is limiting [78, 213, 214]. Sulfopeptide mimetics of the tyrosine-sulfated CCR5 amino-terminus also bind gp120 and inhibit HIV-1 entry [112, 148]. These peptides bind inefficiently in the absence of CD4, but more efficiently when gp120 is bound by CD4 or a CD4-mimetic peptide [78, 112, 148]. We reasoned that fusions of a CD4 and CCR5-mimetic peptides (represented in Fig. 4.1) would bind gp120 cooperatively, and that the CCR5-mimetic peptide would limit the ability of the CD4-mimetic peptide to promote infection.

A

Peptide	Sequence
CCR5mim3	GDYYDYDGGYYYD
CD4mim6	DKLHSCKVLCGSLGLHGTLRGSFCACVVG
1-Fc	GDYYDYDGGYYYDGDKLHSCKVLCGSLGLHGTLRGSFCACVVG
6-Fc	GDYYDYDGGYYYDGGAAGGDKLHSCKVLCGSLGLHGTLRGSFCACVVG
11-Fc	GDYYDYDGGYYYDGGAGGEAGAGGDKLHSCKVLCGSLGLHGTLRGSFCACVVG
16-Fc	GDYYDYDGGYYYDGGAGAGGEGAGAGAGGDKLHSCKVLCGSLGLHGTLRGSFCACVVG
Rev-Fc	DKLHSCKVLCGSLGLHGTLRGSFCACVVGGGAGGEAGAGGGDYYDYDGGYYYD

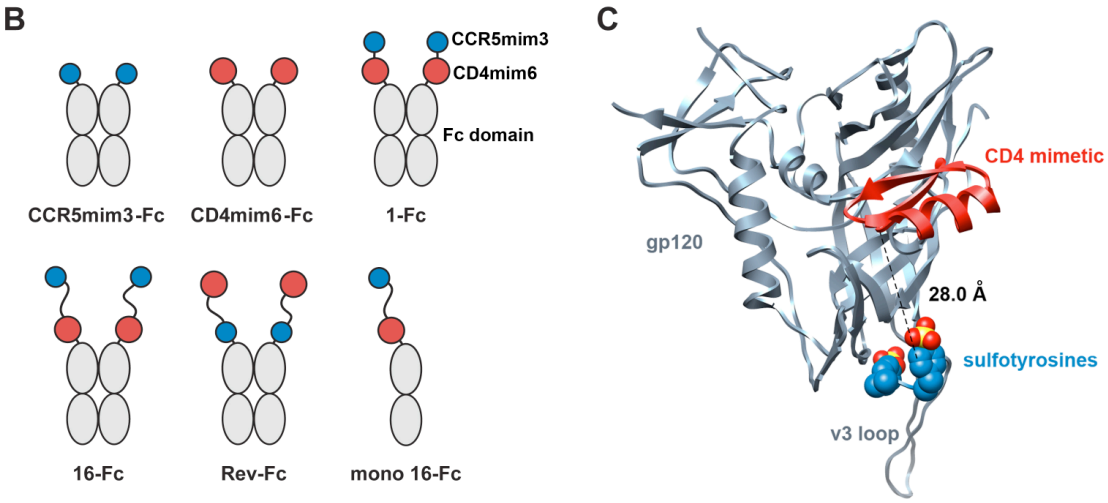

Figure 4.1. Receptor-mimetic constructs. A, Sequences of the single- and double-mimetic peptides used in this work. Blue indicates CCR5mim3 and red indicates CD4mim6. Linker regions are not highlighted. **B,** Diagram of mimetic-Fc fusions constructs. CCR5mim3 is shown as a blue circle, CD4mim6, as a red circle, and the Fc domains are indicated in grey. **C,** Ribbon diagram of HIV-1 gp120 (grey), bound to a CD4-mimetic peptide (red) and sulfotyrosines (blue). Sulfotyrosines derive from the heavy chain of the CD4-inducible antibody 412d. Figure was generated by aligning the gp120 molecules of the gp120-CD4-412d complex (2QAD) with the gp120-F23 complex (1YYM) [62,146]. Dashed line indicates the 28 angstrom distance from the amino-terminal α-carbon of the CD4 mimetic to the α-carbon 412d sulfotyrosine 100.

4. C. METHODS

Plasmids and cells. Double-mimetic peptides and all mutations were created using inverse PCR on a pUC19-derived plasmid carrying the expression cassette from the previously described pCDM8-derived plasmid expressing the signal sequence of CD5 and the Fc domain of human IgG1 [215] using Phusion Flash Mix (Thermo/Finnzymes)

with added sequence and or mutations encoded by phosphorylated primers. Products were ligated using T4 DNA Ligase (NEB) and transformed into DH5α. Plasmids encoding monomeric Fc forms were generated by altering the human IgG1 Fc domain codons to encode the following changes: C219A, C226N, C229G, N297D, L368R, F405H, and Y407E (numbering according to the Kabat database). Plasmids expressing HIV-1 gp120s and gp160s have been described previously [78]. Human embryonic kidney 293T and TZM-bl cell lines were obtained from the American Type Culture Collection.

Protein production. Maxiprep DNA was prepared from 250 mL overnight LB cultures of DH5α carrying the Amp resistant plasmids using PureLink maxiprep kits (Invitrogen) according to the manufacturer's instructions. 293T cells in 140 mm plates were transiently transfected with 25 μg/plate at 50% confluency by standard Ca Phosphate transfection. At 12 hrs post-transfection, 10% FCS-DMEM media was replaced with serum free 293 Freestyle expression media (Invitrogen). Culture supernatants were harvested 48 hrs later, debris was cleared by centrifugation for 10 min at 1,500g and was filtered in 0.45 μm filter flasks (Millipore). Complete protease inhibitor cocktail (Roche) was added to the filtered supernatants. 500 μl bed volume of Protein A sepharose beads (GE Healthcare) were added and were agitated overnight at 4°C. The bead/supernatant mixture was collected by gravity flow column (Biorad) and was washed with 30 mL PBS (Lonza) + 0.5 M NaCl (0.65 M NaCl final) followed by 10 mL PBS. Protein was eluted with 5 mL 2 M arginine pH4 into 1 mL 1 M Tris pH 7.5. Buffer was exchanged for PBS and protein was concentrated to 1 mg/ml by Ultrafiltration (Amicon Ultra) at 3,000g.

HIV-1 neutralization studies. Pseudotyped virus was produced by coexpression of envelope glycoproteins of the indicated HIV isolates with NL43ΔEnv. 293T cells at 50% confluency in T175 (Falcon) flasks were transfected with 25 μg of plasmid encoding envelope glycoprotein and 45 μg of NLΔEnv by the standard Calcium Phosphate technique. 10% FCS DMEM media was changed at 12 h and media was collected at 48 hrs post-transfection. Viral supernatants were cleared by centrifugation for 10 min at 1,500g, passed through a 0.45 μm syringe filter (Millipore), and stored at -80°C. Neutralization by Fc-fusion constructs was performed according to the protocol described in Li et al. [209]. In brief, varying concentrations of purified Fc fusion proteins were incubated with virus at a total volume of 100 μl per well in 96 well plates for one hour after which 10,000 cells/well of TZM-bl cells in 100 μl of media were added. Plates were incubated 72 h at 37°C after which 100 μl of media was replaced with freshly prepared Brite-Lite reagent (Perkin-Elmer), and luminescence data was collected on a Victor3V (Perkin Elmer). All neutralizations were performed in triplicate.

Infection of CCR5-positive, CD4-negative cells. CF2th-CCR5 cells were incubated with pseudoviruses generated with pNL4-3.Luc.R-.E-bearing the indicated HIV-1 envelope glycoprotein and encoding firefly luciferase in the presence of increasing amounts of CD4-mimetic (CD4mim6-Fc) double-mimetic (11-Fc) or CD4-Fc. Infection was measured as in neutralization studies.

Immunoprecipitation. HIV-1 gp120 constructs were transfected into 6-well plates of 50% confluent 293T cells at 1 μg DNA/well by standard calcium phosphate technique. 12 h post-transfection, media was replaced with cysteine/methionine-negative DMEM containing ExpressLabel (Cole Parmer) and 5% dialyzed calf serum (Invitrogen).

Supernatants were collected 24 h later, and were cleared by centrifugation for 2 min at 3,000 rpm, followed by 1 min at 10,000 rpm and complete protease inhibitor cocktail (Roche) was added. Immunoprecipitations were conducted by binding 1 μg of Fc or monomeric Fc constructs to protein A beads for 2 hrs at 4°C with rocking. Beads were then washed twice with PBS-Tween (0.01%) and ^{35}S-radiolabeled gp120 was added and incubated an additional 2 hrs at 4°C with rocking. Beads were washed 3 times with PBS-T and protein was eluted by heating 10 min at 70°C in NuPage sample buffer. Samples were run on NuPage bis-tris gels, treated with methanol/acetic acid fixative and dried on Whatman paper. Radiolabeled protein was quantified with Phosphoimager (Fuji).

Staining of cell expressed HIV-1 envelope glycoprotein. Cells expressing HIV-1 envelope glycoprotein trimers were prepared by transient transfection of plasmids expressing envelope glycoproteins truncated in the C-terminal cytoplasmic domains to facilitate surface expression (gp160-ΔCT). T75 flasks of 293T cells at 50% confluency were transfected with 20 μg of plasmid encoding gp160-ΔCT and 3 μg of plasmid encoding Tat by the calcium phosphate technique and incubated overnight. DMEM + 10% FCS was changed at 18 h post transfection. Cells were collected 36 h post transfection by treatment with non-enzymatic cell dissociation solution (Sigma), followed by centrifugation for 5 m at 1200 rpm. Cells were resuspended in flow cytometry buffer (PBS + 2% Donor Goat Serum + 0.1% sodium azide). Cells were incubated with Fc-fusion proteins diluted to varying concentrations in FACS buffer for 45 min on ice, followed by two FACS buffer washes. APC-conjugated anti-human Fc-gamma (Jackson ImmunoResearch) diluted in FACS buffer was added and cells were incubated on ice for 30 m. Cells were then washed 2 times with flow cytometry buffer then fixed for flow

cytometric analysis with 1% paraformaldehyde in PBS. Flow cytometery data were analyzed using FlowJo.

4. D. RESULTS

Inspection of the structure of gp120 bound to one CD4-mimetic peptide, F23, indicated that its amino-terminal residue was held by a disulfide bond in an orientation that sterically precluded fusion at its amino-terminus [146, 213]. Also, its carboxy-terminus was oriented away from the gp120 sulfotyrosine binding pockets, and likely required a prohibitively long linker to allow for both mimetic peptides to bind simultaneously (Fig. 4.1C). We accordingly modified a natural amino-acid form of this CD4-mimetic peptide to remove its first disulfide bond and permit the addition of amino-terminal linkers [213]. The affinity of this peptide was subsequently improved by phage-display [213].

Using this improved CD4-mimetic peptide, CD4mim6 [213], and CCR5mim3, a shortened form of a previously described CCR5 mimetic peptide [148], we produced a series of CCR5mim3/CD4mim6 fusion proteins separated by linkers of 1, 6, 11, and 16 amino acids (1-Fc, 6-Fc, 11-Fc, 16-Fc; Fig. 4.1A) and a reverse CD4mim6/CCR5mim3 fusion protein with a linker of 11 amino acids (Rev-Fc; Fig. 4.1A-B). The fusion protein with the shortest linker, 1-Fc, was consistently the least efficient at neutralizing HIV-1, and was in most cases less efficient than CD4-Ig (Fig. 4.2A). Fusion proteins with linker lengths of 11 and 16 neutralized all HIV-1 isolates with comparable efficiency, and with markedly greater efficiency than CD4-Ig. With some isolates (HXB2, SG3) 6-Fc, with a

six amino-acid linker, neutralized as efficiently as 11-Fc and 16-Fc. With others (89.6, BaL), 6-Fc was clearly less efficient than 11-Fc. These data suggest that, with linkers of 11 or greater, both mimetic peptides are engaged, likely to the same gp120 monomer. Consistent with this, mixtures of both mimetic peptides alone, each at the same total concentration as 16-Fc, neutralized HIV-1 much less efficiently than 16-Fc (Fig. 4.2B). Similarly, Rev-Fc, a construct in which the order of mimetic peptides was reversed, also neutralized less efficiently than 16-Fc (Fig. 4.2B). Thus 11-Fc and 16-Fc potently neutralize HIV-1, likely because they simultaneously bind the CD4- and coreceptor-binding sites of gp120.

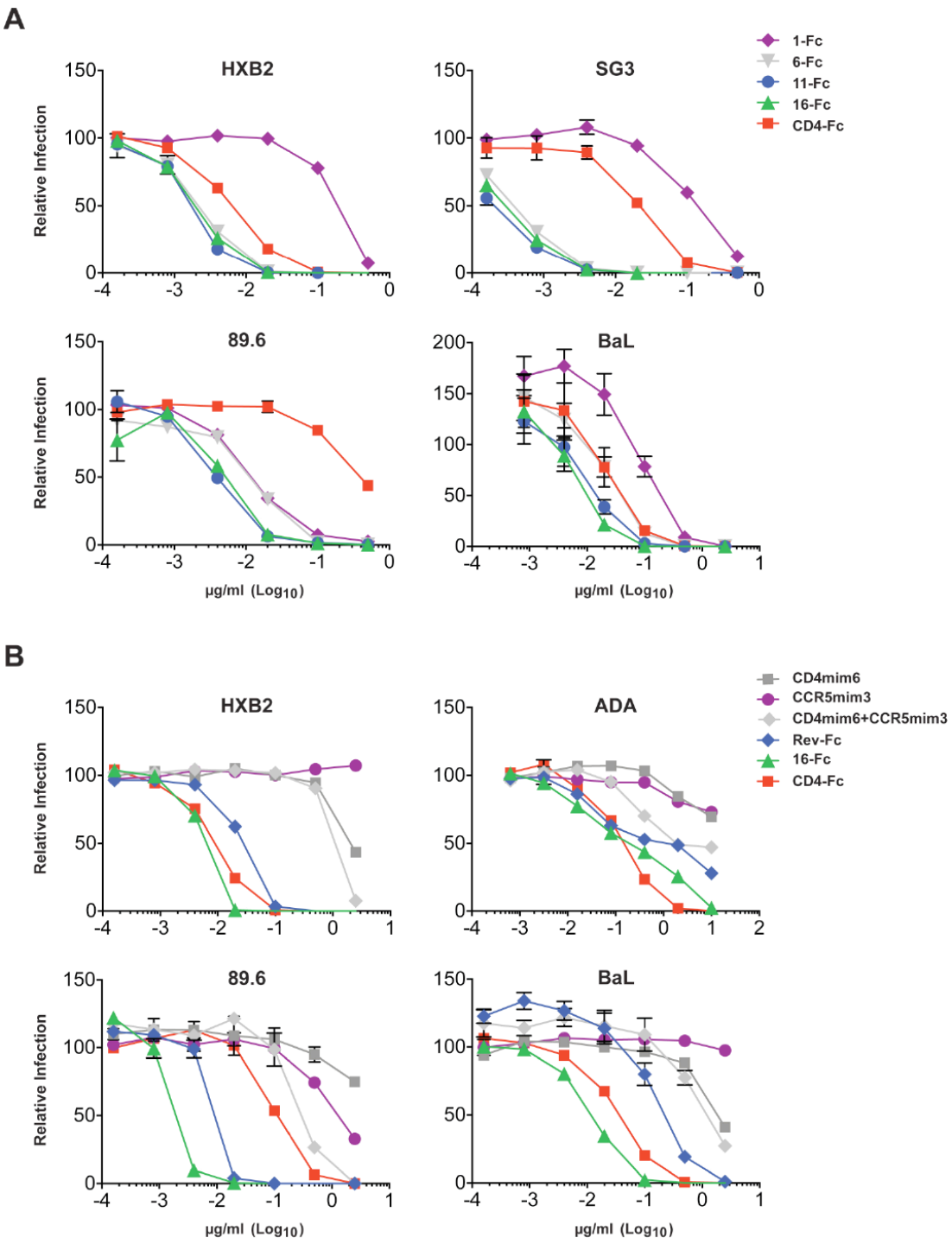

Figure 4.2. Neutralization assays of single- and double-mimetic peptides. A, HIV-1 pseudotyped with the envelope glycoproteins of the indicated HIV-1 isolates were incubated with TZM-bl cells and with CD4-Fc or double-mimetic Fc fusions varying in linker length. Infection was determined by luciferase activity 48 hours later. **B,** Assay similar to that in A except that CD4mim6-Fc, CCR5mim3-Fc, a mixture of these constructs, Rev-Fc, 16-Fc, and CD4-Fc were assayed. All neutralization studies were performed in triplicate, with bars representing standard error. Experiments are representative of two with similar results.

The neutralization activity of soluble CD4 and CD4-mimetic peptides can be attenuated by their parallel tendency to promote infection, possibly by promoting interaction of the HIV-1 envelope glycoprotein with the coreceptor [79, 216]. We therefore sought to determine if CCR5mim3 prevented induction of infection by CD4mim6, explaining in part the potency of 11-Fc. Both CD4-Fc and CD4mim-Fc enhanced HIV-1 infection in cells expressing CCR5 but not CD4, whereas 11-Fc did not (Fig. 4.3). Thus the CCR5mim3 domain of 11-Fc blocks the ability of CD4mim6 to promote infection.

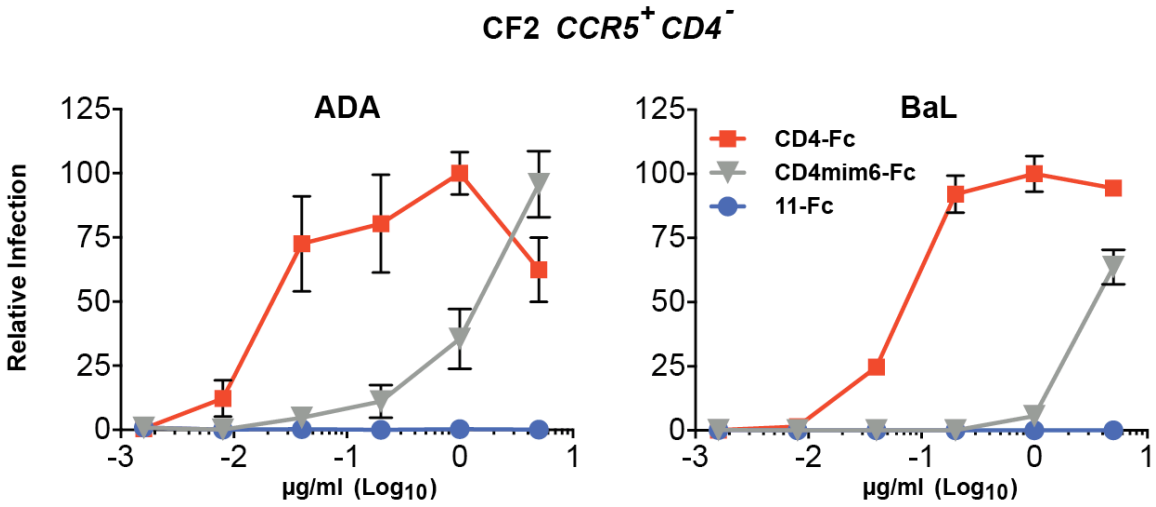

Figure 4.3. Infection of CCR5-positive, CD4-negative cells. CCR5-high, CD4-negative Cf2Th cells were incubated with the indicated pseudoviruses encoding firefly luciferase in the presence of the indicated concentrations of CD4-Fc, CD4mim6-Fc, and 11-Fc. Infection, measured as luciferase activity as in Fig. 2. Numbers represent relative luciferase normalized to largest value. Experiment was performed in triplicate, with bars representing standard error.

Double mimetic peptides with sub-optimal linkers or orientations still neutralize better than mixtures of CCR5mim3-Fc and CD4mim6-Fc (Fig 4.2B), perhaps because gp120 binds CCR5mim3 from one arm of the Fc dimer and CD4mim6 from the opposite arm. To explore this possibility we generated monomeric Fc forms of 1-Fc and 16-Fc

(mono 1-Fc, mono 16-Fc (see Fig. 4.1B) [78]. These variants were compared with dimerized forms of the same constructs by flow cytometry with cells expressing HIV-1 envelope glycoprotein trimers (Fig. 4.4A), in neutralization studies (Fig. 4.4B), and by immunoprecipitation of radiolabeled gp120 (Fig. 4.4C). In each assay, the dimer form of 16-Fc modestly outperformed the monomer form of 16-Fc and dimer form of 1-Fc. All three of these constructs bound envelope glycoprotein trimers, precipitated gp120 and neutralized HIV-1 markedly more efficiently than monomeric 1-Fc. Fig. 4.4D diagrams our interpretations of this data, namely that monomeric and dimeric 16-Fc, and dimeric 1--Fc can bind both receptor sites on gp120, but monomeric 1-Fc can bind only one. Note that the monomeric Fc appears to expose a protease site in the CH3 Fc domain, but that this cleavage does not interfere with its association with Protein A or the binding and neutralization activities of monomeric fusion constructs.

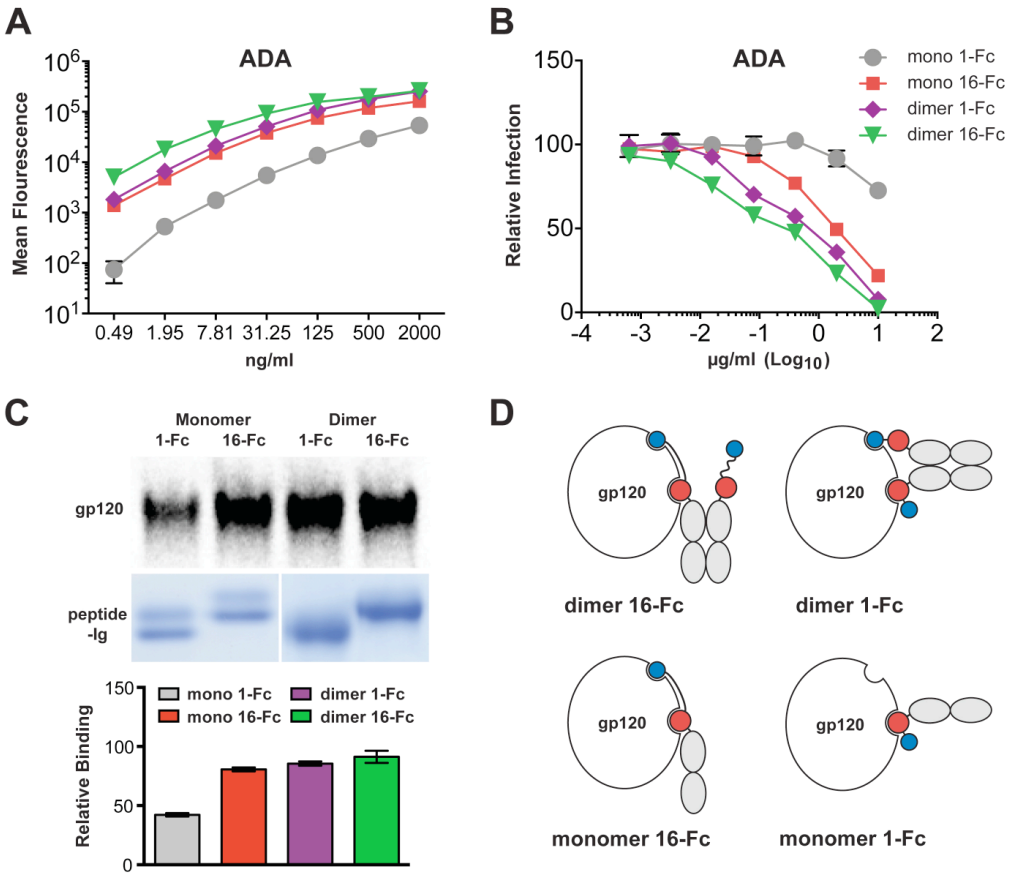

Figure 4.4. Comparison of monomer and dimer forms of double-mimetic peptides. Monomeric and dimeric Fc fusions of the indicated double mimetic peptides were assayed by: **A,** cell-surface binding to 293T cells expressing the envelope glycoprotein of the ADA isolate, as measured by flow cytometry; **B,** neutralization studies, as described for Fig. 2; **C,** precipitation of [^{35}S]-cysteine-labeled gp120 by Fc-fusions bound to Protein A-sepharose beads. **D,** A model explaining the precipitation results in C. Figures A-C were performed in triplicate, with bars indicating standard error, and are representative of at least two experiments with similar results.

4. E. DISCUSSION

Collectively our data describe a novel class of peptide inhibitors of HIV-1 entry, namely those that can simultaneously bind both the receptor- and coreceptor-binding sites of gp120. These peptides are more potent than any single-mimetic peptide that targets gp120, and indeed more potent than CD4-Ig. Interestingly, each component complements a weakness of the other. CCR5-mimetic peptides like CCR5mim3 bind gp120 inefficiently, but their affinity increases in the presence of CD4-mimetic peptides or soluble CD4. CD4-mimetic peptides like CD4mim6 bind more efficiently, but they tend to promote infection when cell-expressed CD4 is limiting. CCR5mim6 prevents this, presumably by blocking access to cell-expressed CCR5. With appropriate linkers, double-mimetic peptides also benefit from the increased avidity by binding two discrete and conserved sites of gp120.

Despite these advantages, these peptides have several limitations as therapeutics. First, the CD4-mimetic component is considerably less broad than soluble CD4 or the best CD4-binding site antibodies [213]. Of course, further broadening of this component through phage-based improvement or perhaps an alternative design is possible. Second, the presence of multiple sulfotyrosines complicates chemical synthesis [217]. However bacterial expression systems have been developed which can incorporate sulfotyrosines during translation [218, 219]. Finally, like any peptide, those of this class cannot be administered orally, and given the wide array of oral therapeutics now available, these peptides at best might be useful as part of a microbicide. However, as this study illustrates, these peptides are useful as probes of the structure and function of the HIV-1 envelope glycoprotein.

CHAPTER 5: DISCUSSION

The current work has highlighted a number of features of HIV-1 Env and its inhibitors, while describing four key technologies. In Chapter 2 we described a novel phage display vector that expresses selected products in mammalian cell culture, greatly streamlining the iterative improvement of immunoadhesins and peptides. In Chapter 3 we demonstrated the utility of this vector by using it to improve a CD4-mimetic peptide, and described the use of this peptide in probing Env structure and function. In Chapter 4 we combined this improved peptide together with a coreceptor-mimetic peptide to generate a double-mimetic peptide more potent than the sum of its parts. Finally, in the process of improving our mimetic peptides, we used controlled diversification techniques with applications beyond phage display.

In the following sections, we discuss how each of these technologies might be extended, and discuss some aspects of Env function that this work has highlighted. First two useful extensions of the pDQ1 vector are described: immediate mammalian expression of whole antibodies derived from phage selection of Fabs, and a single screen combining the strengths of phage and mammalian display technologies. An approach to further enhancing the breadth and potency of the CD4-mimetic is then described, together with a discussion of the tendency of this mimetic to augment infection under certain conditions. This augmentation complicates efforts to substantially increase the potency of this class of inhibitors. The double-mimetic peptide, which solves this problem, is then discussed, as well as some of the issues associated with further improving its potency. Finally, we describe a novel use of the directed diversification techniques with the pDQ1 to improve the CD4-mimetic. Specifically we discuss how it might be applied

to obtaining a comprehensive picture of pathways of viral escape from antibodies inhibitors such as those described here.

The dual-display vector technology

Development of the dual-display vector. The selection of protein variants from phage-display libraries presents particular challenges when expression of the selected variants necessarily involves expression in mammalian cell culture. Thus, it is often necessary to introduce a subcloning step into the process. Also it is particularly convenient at this point to express these variants as fusions to IgG Fc fragments, thereby enabling further characterization of these variants using a range of tools. Due to differential growth and expression artifacts, variants enriched by panning are not necessarily those that best bind when expressed from mammalian cells. Therefore, it is desirable to evaluate as many output variants as possible in mammalian cell culture, as early as possible in the process. The dual-display vector that we have described (pDQ1) allows for direct expression of output variants when transfected in mammalian cell culture, greatly streamlining this process. This, however, is just the beginning of the potential uses of this type of vector. We anticipate two adaptations of pDQ1: for the display of Fabs (which we can then express as full-length IgG in tissue culture), and the useful integration of phage and mammalian surface display systems.

Dual-display of Fab and antibody libraries. The pDQ1 vector can be readily adapted for the display of Fabs. A full-length IgG consists of two Fabs and an Fc fragment. Using this system, outputs from the Fab library could be immediately

expressed as full-length antibodies in mammalian cell culture without subcloning. There are two alternative strategies to accomplishing this end. A fairly simple implementation would be a two-plasmid system, one each for heavy and light chain. Only one chain at a time is diversified, and only that plasmid would bear a phage origin of replication. After an initial round of transformation and helper phage rescue for the diversified chain alone, progeny phage would be used to transduce *E. coli* in which the undiversified chain is being maintained with different antibiotic resistance. For example, plasmid bearing the phage ori and a diversified heavy chain would have ampicillin resistance, and the second, bearing a fixed light chain would have chloramphenicol resistance. Both would have a mammalian expression cassette with the bacterial promoter and signal sequence embedded in the mammalian signal sequence intron, as is pDQ1 (see Fig. 2.1C). However, only the heavy-chain construct would have the pIII fusion, which would be embedded in the IVS1 intron directly before the Fc hinge. Rescue of this library would provide phage that displays a complete Fab, which then would be used for the initial round of panning. If it were desirable to diversify both heavy and light chains, one could do so by alternating rounds of library construction and selection. The same plasmids, selected from phage panning, would be cotransfected for production of full length IgG in mammalian tissue culture.

A second potential implementation would insert both heavy and light chains into the same plasmid. This approach would allow for simultaneous diversification of both chains, but it requires a larger, more complicated construct, and the mammalian expression product would differ slightly from a native IgG. The light chain would come first, followed by a 2A "protease" ribosomal skip sequence, which would include a strong

intron, within which a stop codon for the light-chain Fab would be embedded. This sequence would be followed by a second bacterial promoter, and a signal sequence for the heavy chain. In mammalian cell culture, the heavy chain would differ from native IgG by a proline at its N-terminus. The C-terminus of the light chain would bear the remainder of the 2A sequence, unless a furin cleavage site were placed before it, in which case the majority would be cleaved off, and then trimmed down to native sequence by carboxypeptidase B. The size of this plasmid (6.9 kb) would be on the large side for phage display, and careful handling would be required to avoid overgrowth by truncated phagemids, but this size is well within the packaging capabilities of filamentous phage.

Integrating Phage and Mammalian Display. Mammalian display libraries are constrained by the size of the initial library at the time of construction and by the limited throughput of panning by cell sorting. The size of the initial library is tied to the number of cells at maximum expansion. Many variants will be cotransfected, with the variants per cell declining over time due to asymmetric segregation as the cells divide. The limited throughput is due to the serial rather than parallel nature of cell sorting, where the maximum throughput is determined by events per second of the cell sorting. Limited throughput can be circumvented using magnetic cell sorting in an initial round of selection. The dual display system of the pDQ1 vector could be easily adapted for use as a mammalian display system, allowing the initial magnetic bead selection of a very complex library to be performed in a phage system. The much simplified, partially selected library would then be transferred to mammalian display selection system for later rounds of selection.

The simplest adaptation would add a transmembrane domain, such as the PDGFR-TM used in the pDisplay system, to the C-terminus of the Fc domain. For panning purposes, only a small fraction of the secreted Fc needs to be expressed at the surface. It is possible to retain use of the soluble variant-Fc fusions by implementing an alternative splice system such as seen in Horlick et al. [178]. In this system, an inefficient splice acceptor removes a stop codon containing sequence, allowing readthrough to a transmembrane domain on a minority of mRNA scripts. Another alternative uses inducible readthrough translation: a leaky stop codon would be placed between the Fc and the transmembrane domain. An aminoglycoside such as G418, is then added 24 hours before selection to promote the surface expression of the POI. Aminoglycosides induce readthrough of stop codons in certain contexts, and such an system has been used to control expression [220].

An improved CD4-mimetic peptide

Development of the CD4-mimetic peptide. Our first application of the dual-display vector was to improve a CD4-mimetic peptide. Our longer-term goal with this effort was the creation of a double-mimetic peptide. Structural analyses suggested that an optimal peptide would initiate with a tyrosine-sulfated CCR5-mimetic peptide fused to the N-terminus of the CD4-mimetic peptide. As a starting point for adaptation and improvement, we used a CD4-mimetic peptide with three disulfide bonds, the first of which constrains the N-terminus (CD4mim1 in Fig. 3.1B). Elimination of this first disulfide bond gave a weakly active mimetic that served as the starting point for

diversification and display (CD4mim2 in Fig. 3.1B). Initial rounds of panning selected a two-disulfide CD4-mimetic with binding properties against clade B envelopes that were comparable to those of the three-disulfide version (CD4mim4 in Fig. 3.1B). Unexpectedly, and despite the fact that selection was against ADA, a clade B Env, binding to clade C Envs was vastly improved over the starting three-disulfide version. We speculate that the accidental improvement in breadth was due to increased flexibility of the mimetic with only two disulfide bonds. An additional iteration through the process yielded a greatly improved CD4-mimetic peptide (CD4mim6 in Fig. 3.1B). We characterized this mimetic peptide for binding and neutralization properties. As described in Chapter 4, we then used this peptide in the creation of double-mimetic peptides (Fig. 4.1A). These studies highlighted a critical weakness of CD4-like inhibitors, namely, their parallel tendency to promote infection under some conditions.

Augmentation of infection by soluble CD4 and CD4-mimetic. CD4 acts both as a primary receptor for HIV-1 and as an attachment factor. Soluble CD4 (sCD4) is able to enhance infection when cell-surface CD4 is limiting, or complement the absence of CD4 in a coreceptor-positive cell line [216, 221, 222]. When CD4 is not limiting, sCD4 interferes with the ability of cellular CD4 to concentrate virus at the cell surface for interaction with coreceptor, but it can still enhance infection in cell culture at subneutralizing concentrations. The enhancement at low concentrations (compared to the high concentrations necessary for receptor complementation) suggests a trimer occupancy model, where occupation of one gp120 induces changes in the trimer as a whole, reducing the entropic barrier to the binding not only of coreceptor but also perhaps of other CD4 molecules. Above a certain concentration, multiple occupancy of

the trimer by sCD4 molecules is favored and neutralization overcomes enhancement. Stated another way, CD4 may induce its own binding, and certainly promotes coreceptor association.

The thermodynamic data we have in support of the conformational masking hypothesis all derive from ITC measurements of gp120 binding. Careful consideration of sCD4 enhancement in light of the global rearrangements of various liganded and unliganded states observed by Tran et al. suggest a more complicated interplay between tertiary and quaternary rearrangements. The difference between primary isolates and lab-adapted isolates in CD4-binding is much greater for trimeric Env than for soluble gp120 [101, 223]. Perhaps there is an additional energy barrier associated with quaternary shifts in trimeric Env that is paid by initial binding of one subunit. For example, the CD4-bound tertiary structure may be induced in the other gp120 subunits of the trimer when CD4 binds a single subunit.

Like soluble CD4, CD4-mimetic peptides are able to enhance infection and complement the absence of cellular CD4. Also like sCD4, enhancement of infection in cell culture is observed at sub-neutralizing concentrations (Fig. 4.3). Induction of the CD4-bound conformation appears to require considerably less contact area than provided by CD4. In the crystal structure of the parental CD4-mimetic F23 bound to gp120, the adoption of the CD4-bound conformation is striking. Nonetheless, in light of the more recent structures of the unliganded extended core of gp120, which also adopts the CD4-bound conformation, the significance of the previous structure is less clear. However, receptor complementation and low concentration enhancement are clear indications that our CD4-mimetic peptide induces the CD4-bound state (Figs 3.6A and 4.3). Perhaps

because of the lower affinity, there is a broader more noticeable plateau to this enhancement (Fig. 4.3). Unlike CD4, CD4-mimetic contact with the V1/V2 stem is not directly observed, thus any fixation of the bridging sheet would likely have to be communicated to topological layer 2 via the α-helix in layer 3 contacts with CD4 Phe43, as per the model of Désormeaux el al. [224]. This group found that mutations in layer 3 at its interface with layer 2 had moderate effects on CD4-binding while abolishing the binding of CD4i antibodies. This model suggests a means of fixing the layer 2 that does not require interaction with the V1/V2 stem.

Improvements to the CD4-mimetic. The CD4-mimetic peptide is currently of insufficient breadth to be considered a viable therapeutic and would benefit from maturation against a broader set of Envs, with particular attention paid to the difficult-to-neutralize tier III envelope glycoproteins. Currently the linker region of the double-mimetic peptide is designed to serve only as a flexible, soluble tether between the two mimetic peptides. However, it necessarily traverses the area of CD4 contact with the V1/V2 stem, the a1 α-helix [225, 226]. Residues of this linker region of the double-mimetic peptide could be added to the N-terminus of the CD4-mimetic and then soft-randomized. This would create opportunities for creation of useful contacts in a conserved region.

Another potentially useful alteration in the CD4-mimetic would be elimination of an additional disulfide bond. Synthesis of disulfide-cyclized peptides like the CD4-mimetic increases geometrically in difficulty (and expense) as a second and third disulfide bonds are added. Significantly, the deprotection steps necessary for the sequential formation of the second and third disulfide bonds are incompatible with the

retention of sulfotyrosines, which are labile under acidic conditions. The first disulfide bond requires no special chemistry whatsoever, and is therefore compatible with the synthesis of a tyrosine-sulfated peptide, as in the double-mimetic peptide. We observed gains in breadth associated with the elimination of the first disulfide bond. It would be interesting to observe if a similar gain would be attainable with the elimination of a second, or if the entropy penalty for binding the less restricted peptide would be too expensive. We have no indication at this point whether the C6-C24 disulfide (which constrains the α-helix) or the disulfide C10-C26 (which constrains the β-hairpin) would be better tolerated so libraries for each would be constructed.

The double-mimetic peptide

Block to augmentation by double mimetic. Coreceptor-binding site antibodies are poorly neutralizing because steric interactions with the target cell membrane prevent cooperative binding between cell-surface CD4 and CD4i antibodies. The CCR5-mimetic peptide used here is not subject to this limitation, but it is a weak binder by itself. The CD4-mimetic peptide binds Env more efficiently, but can enhance infection at sub-neutralizing concentrations. We found that addition of a CCR5-mimetic to a CD4-mimetic generated a double-mimetic of greater potency than that of either component alone or as a mixture. Examination of the crystal structures of gp120 bound to a CD4-mimetic peptide, and of gp120 bound to the tyrosine-sulfated antibody 412d, led us to believe that a CCR5-mimetic CD4-mimetic orientation would be optimal [62, 146]. Indeed in Chapter 4, we proved that to be the case. However, our analysis was

complicated by the multiple possible binding modes of what is essentially a homo-dimer (the Fc fusion) of a hetero-dimer (the double-mimetic) binding a trimeric Env target at two distinct sites. To address these issues, we created the first monomeric Fc to appear in the literature, which appeared first in Kwong JA et al. [78] (See Fig. 5.1).

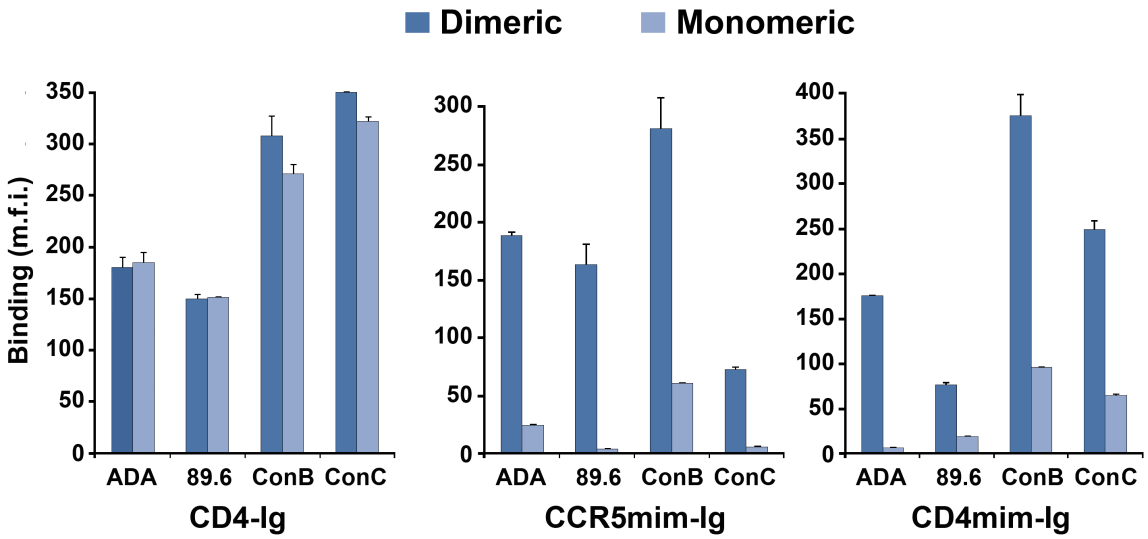

Figure 5.1. CCR5mim-Ig but not CD4-Ig binds two monomers of the HIV-1 envelope glycoprotein trimer. CD4-Ig, CCR5mim-Ig, CD4mim-Ig, or variants thereof which express as monomeric proteins were incubated at approximately half-maximal concentrations (5 nM for CD4-Ig, 50 nM for CD4mim-Ig and CCR5mim-Ig) with 293T cells expressing the indicated envelope glycoproteins and analyzed by flow cytometry. Differences between monomeric and dimeric forms of CCR5mim-Ig and CD4mim-Ig in all cases are significant ($P<0.005$) by Student's t. Adapted from Fig. 2C of Kwong JA et al. [78] using a monomeric Fc domain generated for the studies of Chapter 4

Cis and Cis/Trans binding modes. The double-mimetic peptide is comprised of two moieties that bind cooperatively to two different sites on Env. If the binding partner is monomeric gp120, as is the case for IP and SPR assays, there are two different potential binding modes, where the receptor- and coreceptor-binding sites are occupied by the components of two different molecules of the double-mimetic binding cooperatively (so called *trans* binding, see Fig. 4.4D), or by the two components of the

same molecule (*cis* binding). The situation is complicated further when the double-mimetic peptide is expressed as a fusion to an IgG1 Fc domain, which is dimeric. In this case, binding can potentially involve the CD4-mimetic from one half of the homodimer, and the CCR5-mimetic from the other (*cis*/*trans* dimer binding). We distinguish between these modes of binding using a monomeric Fc version of the double-mimetic as a control. Using these monomers, we observed differences in binding between optimal and sub-optimal linker lengths, indicating that *cis* binding was only achieved with longer linkers.

Yet another layer of complexity is introduced when the binding partner is trimeric Env, as is the case for neutralization or surface-staining assays. Here a homodimer (double mimetic) of a heterodimer (Fc) is binding a homo-trimer (Env), and binding may take place with either component of two halves of the Fc binding across the trimer. Differentiation between these binding modes is made by inference: monomeric and dimeric forms of the single-mimetic components are compared by surface staining assay. This evaluation was performed in Kwong JA et al. [78] on earlier versions of the mimetic peptides and with CD4-Ig (see Fig. 5.1). Those data indicate that the two halves of CD4-Ig could not bind across the trimer, that CD4-mimetic could do so to some extent, but less so than CCR5-mimetic. These data are in good agreement with what can be inferred from cryo-EM data. CD4 binds in the trimer cleft, and projects away from the trimer axis in the lateral plane, placing its C-terminus, and its Fc dimer partner far from the other two clefts of the trimer. The C-terminus of CD4 mimetic is much closer to gp120, and two flexible linkers, and the free portion of the two Fc hinges, might potentially bridge the gap (See Fig. 4.1). In contrast, the coreceptor-binding sites are grouped at the apex of the trimer spike, near the trimer axis of symmetry, easily allowing binding across the trimer.

Improvements to the double mimetic peptide. With adaptation of the pDQ1 vector to mammalian surface expression, it would also be possible to direct evolution of the double-mimetic using this vector. The CCR5-mimetic, the linker region and the N-terminal α-helix of the CD4-mimetic would be subject to soft randomization and an initial round of selection in phage would select on CD4-binding capacity alone, as no tyrosines would be sulfated in this system. Subsequent rounds of selection would be in a mammalian surface display system, with sulfotransferase co-expression. Selection would employ cell-sorting with fluorescently labeled gp120.

Future of the CCR5-mimetic peptide. Without significant improvement, CD4-mimetic based compounds are unlikely to be useful in the clinic. CD4mim6 is probably adequate for most non-clinical use of mimetic peptide as a tool. On the other hand, it has been demonstrated in our lab that a CD4-Fc with a CCR5 element fused to it is a very potent immunoadhesin with an exceptionally broad range of neutralization across all clades and tiers of HIV-1 and even HIV-2 and SIV (Gardner et al., submitted). Thus improvement of the CCR5-mimetic peptide is of greater relevance.

One could improve this fusion directly with a mammalian display system. The CCR5-mimetic peptide of this potent CD4-Fc-CCR5mim construct is fused to the C-terminus of Fc. To evolve this peptide in the proper context one would want to display it on the cell-surface with an N-terminal rather than C-terminal membrane anchor, in a type II membrane protein configuration. The topological morphologies of type I and type II membrane proteins are quite similar, the only difference is that a type II membrane protein lacks a signal-peptidase cleavage site in what is then the signal anchor (the type II trans-membrane domain), and, unlike type I transmembrane domains, there is no stop

anchor at the C-terminus. Otherwise, the orientation of engagement with the signal recognition particle, and the translocation of the growing peptide across the membrane are identical. The N-terminus of CD4 enters domain one (D1) near the junction with domain two (D2), away from the site of gp120 binding. A type II signal anchor (such as from the transferrin receptor) followed by a sufficiently long anchor should allow for display of a CD4-Fc construct with unconstrained binding of an N-terminally anchored CD4, followed by the Fc and a free C-terminal CCR5-mimetic peptide. Diversification would be focused on the CCR5-mimetic element. To emulate genuine trimer binding, fluorescently labeled virus would be used as bait, and selection would be by cell sorting on the fluorophore.

However, given the great difference in affinity between the CCR5-mimetic peptide and sCD4, the double-mimetic peptide might be the preferred platform for directed evolution of CCR5-mimetic peptide. Variation would be restricted to the CCR5-mimetic component, and selection would be by cell surface binding. An advantage to the double-mimetic in this context is that the CCR5-mimetic is N-terminal, thus any library components containing frameshifts could be removed through an initial round of phage display, with subsequent rounds of mammalian display selecting for differences in binding between variants. Alternatively, the CCR5-mimetic could be improved independently by mammalian display with induction by sCD4 provided in *trans* during selection.

A system exists for the display of sulfotyrosines in phage using a sulfotyrosine amber suppressor codon [219, 227]. This approach could indeed be used for evolution of CCR5-mimetic peptides. However, because the sulfotyrosines are hard-coded, there is

no guarantee that outputs would correspond to sequence that would be sulfated in a mammalian expression system. One possible way around this difficulty might be to bias towards retention of sulfation motifs by excluding hydrophobic or positively charges amino acids, though this would necessarily involve the exclusion or reduction of glutamic acid and glycine; both of these amino acids compatible with, and in the case of glycine critical for, sulfation. The dual-display pDQ1 system could not be used with this system because amber-encoded sulfotyrosines would translate as stop codons when transferred to the mammalian system.

Modeling escape pathways

The current work has concentrated on the development of HIV-1 entry inhibitors, and points the way towards directed evolution of other, potentially more potent immunoadhesins. However, an equally interesting and relatively unexplored application of these controlled diversification and selection techniques would be to model the viral side of the conflict.

Directed Evolution of Env. The HIV-1 virus is adept at adapting itself to the changing immune environment of the host over the course of infection. In longitudinal studies of virus and antibody populations, contemporaneous virus is resistant to neutralization whereas virus at earlier time points is sensitive [228]. Adaptation to broadly neutralizing antibodies comes at a fitness cost, and virus resistant to CD4bs bNAbs replicates poorly due to a reduced ability to bind CD4 [229]. In the development of antibody-based therapeutic approaches, it is important to be able to identify cocktails

of antibodies to which, in combination, would close off all avenues of escape [230]. Such *in vitro* checkmate analysis could save years of study in an expensive animal model.

As a model for viral escape, simple passage in cell culture in the presence of bNAbs is inadequate. Passage in cell-culture leads to non-physiologically relevant changes in Env including truncations in V1/V2, and other changes that lead to a open Env trimer conformation and less masking of epitopes [74, 129, 216, 231, 232]. Moreover, typically only a single mutation emerges to overgrow a culture under selection. Also, the endogenous diversification of the virus is inadequate to a compressed time-scale. The fidelity of the viral polymerase is tuned to balance the need to generate viable progeny with the need to generate the diversity necessary for rapid adaptation. Introduction of a higher mutation rate, for example by the restriction factor APOBEC3G, pushes this balance towards catastrophe [233]. Furthermore, mutation due to polymerase infidelity gives no preference to potentially relevant mutations. Mutations in viral proteins other than Env are much more likely to be deleterious, and with no potential to further escape from bNAbs. In contrast, with directed *in vitro* diversification model, non-random mutations can be tolerated at a significantly higher rate.

By concentrating randomization in the one gene relevant for escape, a controlled diversification technique can be used to explore a large sequence space in a much shorter period than the virus could arrive at on its own. It is useful to further exclude from diversification highly conserved residues of Env, thereby increasing the ratio of potentially adaptive to deleterious mutations. Ideally one would want to construct groups of libraries randomizing unconstrained residues in two different areas of Env, such that every combination of two mutations would be covered. Initial diversity could be

evaluated by next-generation sequencing. Serial passage of libraries at low to moderate MOI would eliminate the free-rider effect caused by co-transfection. These virus libraries would then be screened against individual neutralizing antibodies, or against optimized cocktails of such antibodies, providing accelerated insight into viable escape pathways. Given time, rare events such as deletions and insertions give the virus means towards escape that would fall outside the model. Nevertheless, it should be possible to differentiate between antibodies and antibody cocktail by ease of escape. Application of the model to antibodies and cocktails that are already under study in primates would allow for testing of the predictive value of the system.

Closing Remarks

In the conflict between HIV-1 quasispecies and host antibody repertoire the virus holds the distinct advantage. Viral diversification is faster than somatic hypermutation, and viral replication outpaces clonal expansion. The virus provides very few non-moving targets, and distracts immune recognition away from these conserved sites. When recognition does occur, such as in the subset (10-30%) of patients who develop bNAbs in their repertoire, it is not clear that this can happen in time to significantly affect disease progression [98, 234, 235]. However, these bNAbs do indicate points of vulnerability for the virus and serve as starting points for therapeutic development. The broadest antibodies identified so far are directed against the CD4-binding site. Our experience in constructing the double-mimetic peptide highlights the gains to be had in a coordinated attack on the CD4- and coreceptor-binding sites, and demonstrate the potential utility of a

peptide targeting the coreceptor-binding site, despite its low activity by itself. In our experience with directed evolution of the CD4-mimetic peptide, we are struck by the robustness of the diversification techniques and display tools that we have adopted and improved. We expect these tools to be of use in developing a new generation of antibodies and immunoadhesin cocktails. In parallel, analogous directed diversification techniques applied to Env may assist in checkmate analysis of such agents in blocking viral evolution and escape.

REFERENCES

1. Cowley, A.C., N.L. Fuller, R.P. Rand, and V.A. Parsegian, Measurement of repulsive forces between charged phospholipid bilayers. *Biochemistry*, 1978. **17**(15): p. 3163-3168.

2. Geijtenbeek, T.B., D.S. Kwon, R. Torensma, S.J. van Vliet, G.C. van Duijnhoven, J. Middel, . . . Y. van Kooyk, DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell*, 2000. **100**(5): p. 587-597.

3. Arthos, J., C. Cicala, E. Martinelli, K. Macleod, D. Van Ryk, D. Wei, . . . A.S. Fauci, HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells. *Nat. Immunol.*, 2008. **9**(3): p. 301-309.

4. Saphire, A.C., M.D. Bobardt, Z. Zhang, G. David, and P.A. Gallay, Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages. *J Virol*, 2001. **75**(19): p. 9187-9200.

5. Maddon, P.J., A.G. Dalgleish, J.S. McDougal, P.R. Clapham, R.A. Weiss, and R. Axel, The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell*, 1986. **47**(3): p. 333-348.

6. McDougal, J.S., M.S. Kennedy, J.M. Sligh, S.P. Cort, A. Mawle, and J.K. Nicholson, Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science*, 1986. **231**(4736): p. 382-385.

7. Sattentau, Q.J. and J.P. Moore, Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J. Exp. Med.*, 1991. **174**(2): p. 407-415.

8. Thali, M., J.P. Moore, C. Furman, M. Charles, D.D. Ho, J. Robinson, and J. Sodroski, Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J Virol*, 1993. **67**(7): p. 3978-3988.

9. Sattentau, Q.J., J.P. Moore, F. Vignaux, F. Traincard, and P. Poignard, Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J Virol*, 1993. **67**(12): p. 7383-7393.

10. Feng, Y., C.C. Broder, P.E. Kennedy, and E.A. Berger, HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*, 1996. **272**(5263): p. 872-877.

11. Choe, H., M. Farzan, Y. Sun, N. Sullivan, B. Rollins, P.D. Ponath, . . . J. Sodroski, The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*, 1996. **85**(7): p. 1135-48.

12. Deng, H., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, . . . N.R. Landau, Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, 1996. **381**(6584): p. 661-666.

13. Dragic, T., V. Litwin, G.P. Allaway, S.R. Martin, Y. Huang, K.A. Nagashima, . . . W.A. Paxton, HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*, 1996. **381**(6584): p. 667-673.

14. Doranz, B.J., J. Rucker, Y. Yi, R.J. Smyth, M. Samson, S.C. Peiper, . . . R.W. Doms, A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell*, 1996. **85**(7): p. 1149-1158.

15. Chan, D.C., D. Fass, J.M. Berger, and P.S. Kim, Core structure of gp41 from the HIV envelope glycoprotein. *Cell*, 1997. **89**(2): p. 263-273.

16. Harrison, S.C., Mechanism of Membrane Fusion by Viral Envelope Proteins. *Advances in virus research*, 2005. **64**: p. 231-261.

17. Carr, C.M. and P.S. Kim, A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell*, 1993. **73**(4): p. 823-832.

18. Hulme, A.E., O. Perez, and T.J. Hope, Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proceedings of the National Academy of Sciences*, 2011. **108**(24): p. 9975-9980.

19. Yang, Y., T. Fricke, and F. Diaz-Griffero, Inhibition of reverse transcriptase activity increases stability of the HIV-1 core. *J Virol*, 2013. **87**(1): p. 683-687.

20. Gilboa, E., S.W. Mitra, S. Goff, and D. Baltimore, A detailed model of reverse transcription and tests of crucial aspects. *Cell*, 1979. **18**(1): p. 93-100.

21. Hu, W.-S. and S.H. Hughes, HIV-1 reverse transcription. *Cold Spring Harb Perspect Med*, 2012. **2**(10).

22. McDonald, D., Visualization of the intracellular behavior of HIV in living cells. *J Cell Biol*, 2002. **159**(3): p. 441-452.

23. Lewis, P.F. and M. Emerman, Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J Virol*, 1994. **68**(1): p. 510-6.

24. Bukrinsky, M.I., N. Sharova, M.P. Dempsey, T.L. Stanwick, A.G. Bukrinskaya, S. Haggerty, and M. Stevenson, Active nuclear import of human

immunodeficiency virus type 1 preintegration complexes. *Proc Natl Acad Sci USA*, 1992. **89**(14): p. 6580-6584.

25. Katz, R.A., J.G. Greger, P. Boimel, and A.M. Skalka, Human immunodeficiency virus type 1 DNA nuclear import and integration are mitosis independent in cycling cells. *J Virol*, 2003. **77**(24): p. 13412-13417.

26. Lee, K., Z. Ambrose, T.D. Martin, I. Oztop, A. Mulky, J.G. Julias, . . . V.N. KewalRamani, Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe*, 2010. **7**(3): p. 221-233.

27. Cherepanov, P., HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells. *Journal of Biological Chemistry*, 2002. **278**(1): p. 372-381.

28. Maertens, G., LEDGF/p75 Is Essential for Nuclear and Chromosomal Targeting of HIV-1 Integrase in Human Cells. *J Biol Chem*, 2003. **278**(35): p. 33528-33539.

29. Ciuffi, A., M. Llano, E. Poeschla, C. Hoffmann, J. Leipzig, P. Shinn, . . . F. Bushman, A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med*, 2005. **11**(12): p. 1287-1289.

30. Engelman, A., K. Mizuuchi, and R. Craigie, HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, 1991. **67**(6): p. 1211-1221.

31. Miller, M.D., B. Wang, and F.D. Bushman, Human immunodeficiency virus type 1 preintegration complexes containing discontinuous plus strands are competent to integrate in vitro. *J Virol*, 1995. **69**(6): p. 3938-3944.

32. Brin, E., J. Yi, A.M. Skalka, and J. Leis, Modeling the late steps in HIV-1 retroviral integrase-catalyzed DNA integration. *J Biol Chem*, 2000. **275**(50): p. 39287-39295.

33. Yoder, K.E. and F.D. Bushman, Repair of gaps in retroviral DNA integration intermediates. *J Virol*, 2000. **74**(23): p. 11191-11200.

34. Klaver, B. and B. Berkhout, Comparison of 5' and 3' long terminal repeat promoter function in human immunodeficiency virus. *J Virol*, 1994. **68**(6): p. 3830-3840.

35. Sodroski, J., C. Rosen, F. Wong-Staal, S.Z. Salahuddin, M. Popovic, S. Arya, . . . W.A. Haseltine, Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. *Science*, 1985. **227**(4683): p. 171-173.

36. Sodroski, J., W.C. Goh, C. Rosen, A. Dayton, E. Terwilliger, and W. Haseltine, A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature*, 1986. **321**(6068): p. 412-417.

37. Hadzopoulou-Cladaras, M., B.K. Felber, C. Cladaras, A. Athanassopoulos, A. Tse, and G.N. Pavlakis, The rev (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a cis-acting sequence in the env region. *J Virol*, 1989. **63**(3): p. 1265-1274.

38. Felber, B.K., M. Hadzopoulou-Cladaras, C. Cladaras, T. Copeland, and G.N. Pavlakis, rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc Natl Acad Sci USA*, 1989. **86**(5): p. 1495-1499.

39. Frankel, A.D. and J.A. Young, HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.*, 1998. **67**: p. 1-25.

40. Jacks, T., M.D. Power, F.R. Masiarz, P.A. Luciw, P.J. Barr, and H.E. Varmus, Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, 1988. **331**(6153): p. 280-283.

41. Wilson, W., M. Braddock, S.E. Adams, P.D. Rathjen, S.M. Kingsman, and A.J. Kingsman, HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. *Cell*, 1988. **55**(6): p. 1159-1169.

42. Göttlinger, H.G., J.G. Sodroski, and W.A. Haseltine, Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A*, 1989. **86**(15): p. 5781-5.

43. Bryant, M. and L. Ratner, Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc Natl Acad Sci U S A*, 1990. **87**(2): p. 523-7.

44. Campbell, S. and V.M. Vogt, Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. *J Virol*, 1995. **69**(10): p. 6487-97.

45. De Guzman, R.N., Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3 -RNA Recognition Element. *Science*, 1998. **279**(5349): p. 384-388.

46. Göttlinger, H.G., T.T. Dorfman, J.G.J. Sodroski, and W.A.W. Haseltine, Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci USA*, 1991. **88**(8): p. 3195-3199.

47. Peng, C.C., B.K.B. Ho, T.W.T. Chang, and N.T.N. Chang, Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. *J Virol*, 1989. **63**(6): p. 2550-2556.

48. Leonard, C.K., M.W. Spellman, L. Riddle, R.J. Harris, J.N. Thomas, and T.J. Gregory, Assignment of intrachain disulfide bonds and characterization of

potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem*, 1990. **265**(18): p. 10373-10382.

49. Johnson, W.E., J.M. Sauvron, and R.C. Desrosiers, Conserved, N-linked carbohydrates of human immunodeficiency virus type 1 gp41 are largely dispensable for viral replication. *J Virol*, 2001. **75**(23): p. 11426-11436.

50. Li, Y., J.J. Bergeron, L. Luo, W.J. Ou, D.Y. Thomas, and C.Y. Kang, Effects of inefficient cleavage of the signal sequence of HIV-1 gp 120 on its association with calnexin, folding, and intracellular transport. *Proc Natl Acad Sci USA*, 1996. **93**(18): p. 9606-9611.

51. Li, Y., L. Luo, D.Y. Thomas, and C.Y. Kang, The HIV-1 Env protein signal sequence retards its cleavage and down-regulates the glycoprotein folding. *Virology*, 2000. **272**(2): p. 417-428.

52. Land, A., D. Zonneveld, and I. Braakman, Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *FASEB J.*, 2003. **17**(9): p. 1058-1067.

53. Allan, J.S., J.E. Coligan, F. Barin, M.F. McLane, J.G. Sodroski, C.A. Rosen, . . . M. Essex, Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science*, 1985. **228**(4703): p. 1091-1094.

54. Robey, W.G., B. Safai, S. Oroszlan, L.O. Arthur, M.A. Gonda, R.C. Gallo, and P.J. Fischinger, Characterization of envelope and core structural gene products of HTLV-III with sera from AIDS patients. *Science*, 1985. **228**(4699): p. 593-595.

55. Veronese, F.D., A.L. DeVico, T.D. Copeland, S. Oroszlan, R.C. Gallo, and M.G. Sarngadharan, Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. *Science*, 1985. **229**(4720): p. 1402-1405.

56. Helseth, E., U. Olshevsky, C. Furman, and J. Sodroski, Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J Virol*, 1991. **65**(4): p. 2119-2123.

57. Kwong, P.D., R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, and W.A. Hendrickson, Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, 1998. **393**(6686): p. 648-659.

58. Haim, H., I. Salas, and J. Sodroski, Proteolytic processing of the human immunodeficiency virus envelope glycoprotein precursor decreases conformational flexibility. *J Virol*, 2013. **87**(3): p. 1884-1889.

59. Starcich, B.R., B.H. Hahn, G.M. Shaw, P.D. McNeely, S. Modrow, H. Wolf, . . . R.C. Gallo, Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell*, 1986. **45**(5): p. 637-648.

60. Wyatt, R., P.D. Kwong, E. Desjardins, R.W. Sweet, J. Robinson, W.A. Hendrickson, and J.G. Sodroski, The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature*, 1998. **393**(6686): p. 705-711.

61. Wyatt, R. and J. Sodroski, The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science*, 1998. **280**(5371): p. 1884-1888.

62. Huang, C.-c., S.N. Lam, P. Acharya, M. Tang, S.-H. Xiang, S.S.-U. Hussan, . . . P.D. Kwong, Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science*, 2007. **317**(5846): p. 1930-1934.

63. Pancera, M., S. Majeed, Y.-E.A. Ban, L. Chen, C.-c. Huang, L. Kong, . . . P.D. Kwong, Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. 2010.

64. Finzi, A., S.-H. Xiang, B. Pacheco, L. Wang, J. Haight, A. Kassa, . . . J. Sodroski, Topological layers in the HIV-1 gp120 inner domain regulate gp41 interaction and CD4-triggered conformational transitions. *Mol Cell*, 2010. **37**(5): p. 656-667.

65. Guttman, M., M. Kahn, N.K. Garcia, S.-L. Hu, and K.K. Lee, Solution structure, conformational dynamics, and CD4-induced activation in full-length, glycosylated, monomeric HIV gp120. *J Virol*, 2012. **86**(16): p. 8750-8764.

66. Kong, L., C.-c. Huang, S.J. Coales, K.S. Molnar, J. Skinner, Y. Hamuro, and P.D. Kwong, Local conformational stability of HIV-1 gp120 in unliganded and CD4-bound states as defined by amide hydrogen/deuterium exchange. *J Virol*, 2010. **84**(19): p. 10311-10321.

67. Kwong, P.D., M.L. Doyle, D.J. Casper, C. Cicala, S.A. Leavitt, S. Majeed, . . . J. Arthos, HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature*, 2002. **420**(6916): p. 678-682.

68. Chen, B., E.M. Vogan, H. Gong, J.J. Skehel, D.C. Wiley, and S.C. Harrison, Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature*, 2005. **433**(7028): p. 834-841.

69. Chen, L., Y.D. Kwon, T. Zhou, X. Wu, S. O'Dell, L. Cavacini, . . . P.D. Kwong, Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science*, 2009. **326**(5956): p. 1123-1127.

70. Zhou, T., L. Xu, B. Dey, A.J. Hessell, D. Van Ryk, S.-H. Xiang, . . . P.D. Kwong, Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature*, 2007. **445**(7129): p. 732-737.

71. Liu, J., A. Bartesaghi, M.J. Borgnia, G. Sapiro, and S. Subramaniam, Molecular architecture of native HIV-1 gp120 trimers. *Nature*, 2008. **455**(7209): p. 109-113.

72. White, T.A., A. Bartesaghi, M.J. Borgnia, J.R. Meyerson, M.J.V. de la Cruz, J.W. Bess, . . . S. Subramaniam, PLOS Pathogens: Molecular Architectures of Trimeric SIV and HIV-1 Envelope Glycoproteins on Intact Viruses: Strain-Dependent Variation in Quaternary Structure. *PLoS Pathog*, 2010. **6**(12): p. e1001249.

73. Kwon, Y.D., A. Finzi, X. Wu, C. Dogo-Isonagie, L.K. Lee, L.R. Moore, . . . P.D. Kwong, Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. *Proceedings of the National Academy of Sciences*, 2012. **109**(15): p. 5663-5668.

74. Tran, E.E.H., M.J. Borgnia, O. Kuybeda, D.M. Schauder, A. Bartesaghi, G.A. Frank, . . . S. Subramaniam, Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. *PLoS Pathog*, 2012. **8**(7): p. e1002797.

75. Myszka, D.G., R.W. Sweet, P. Hensley, M. Brigham-Burke, P.D. Kwong, W.A. Hendrickson, . . . M.L. Doyle, Energetics of the HIV gp120-CD4 binding reaction. *Proc Natl Acad Sci USA*, 2000. **97**(16): p. 9026-9031.

76. Farzan, M., T. Mirzabekov, P. Kolchinsky, R. Wyatt, M. Cayabyab, N.P. Gerard, . . . H. Choe, Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell*, 1999. **96**(5): p. 667-676.

77. Farzan, M., G.J. Babcock, N. Vasilieva, P.L. Wright, E. Kiprilov, T. Mirzabekov, and H. Choe, The role of post-translational modifications of the CXCR4 amino terminus in stromal-derived factor 1 alpha association and HIV-1 entry. *J Biol Chem*, 2002. **277**(33): p. 29484-29489.

78. Kwong, J.A., T. Dorfman, B.D. Quinlan, J.J. Chiang, A.A. Ahmed, H. Choe, and M. Farzan, A tyrosine-sulfated CCR5-mimetic peptide promotes conformational transitions in the HIV-1 envelope glycoprotein. *J Virol*, 2011. **85**(15): p. 7563-7571.

79. Wu, L., N.P. Gerard, R. Wyatt, H. Choe, C. Paralin, N. Ruffing, . . . J. Sodroski, CD4-induced interactions of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature*, 1996. **184**: p. 179-183.

80. Liu, L., R. Cimbro, P. Lusso, and E.A. Berger, Intraprotomer masking of third variable loop (V3) epitopes by the first and second variable loops (V1V2) within the native HIV-1 envelope glycoprotein trimer. *Proceedings of the National Academy of Sciences*, 2011. **108**(50): p. 20148-20153.

81. Kolchinsky, P., E. Kiprilov, and J. Sodroski, Increased neutralization sensitivity of CD4-independent human immunodeficiency virus variants. *J Virol*, 2001. **75**(5): p. 2041-2050.

82. Labrijn, A.F., P. Poignard, A. Raja, M.B. Zwick, K. Delgado, M. Franti, . . . D.R. Burton, Access of antibody molecules to the conserved coreceptor binding site on glycoprotein gp120 is sterically restricted on primary human immunodeficiency virus type 1. *J Virol*, 2003. **77**(19): p. 10557-10565.

83. Moore, J.P., J.A. McKeating, R.A. Weiss, and Q.J. Sattentau, Dissociation of gp120 from HIV-1 virions induced by soluble CD4. *Science*, 1990. **250**(4984): p. 1139-1142.

84. Furuta, R.A., C.T. Wild, Y. Weng, and C.D. Weiss, Capture of an early fusion-active conformation of HIV-1 gp41. *Nat. Struct. Mol. Biol.*, 1998. **5**(4): p. 276-279.

85. Chan, D.C., C.T. Chutkowski, and P.S. Kim, Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc Natl Acad Sci USA*, 1998. **95**(26): p. 15613-15617.

86. Kahle, K.M., H.K. Steger, and M.J. Root, Asymmetric deactivation of HIV-1 gp41 following fusion inhibitor binding. *PLoS Pathog*, 2009. **5**(11): p. e1000674.

87. Wild, C.T., D.C. Shugars, T.K. Greenwell, C.B. McDanal, and T.J. Matthews, Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc Natl Acad Sci U S A*, 1994. **91**(21): p. 9770-4.

88. Weissenhorn, W., A. Dessen, S.C. Harrison, J.J. Skehel, and D.C. Wiley, Atomic structure of the ectodomain from HIV-1 gp41. *Nature*, 1997. **387**(6631): p. 426-430.

89. Gallo, S.A., C.M. Finnegan, M. Viard, Y. Raviv, A. Dimitrov, S.S. Rawat, . . . R. Blumenthal, The HIV Env-mediated fusion reaction. *Biochim Biophys Acta*, 2003. **1614**(1): p. 36-50.

90. Tomaras, G.D., N.L. Yates, P. Liu, L. Qin, G.G. Fouda, L.L. Chavez, . . . B.F. Haynes, Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J Virol*, 2008. **82**(24): p. 12449-12463.

91. Moore, J.P., Y. Cao, D.D. Ho, and R.A. Koup, Development of the anti-gp120 antibody response during seroconversion to human immunodeficiency virus type 1. *J Virol*, 1994. **68**(8): p. 5142-5155.

92. Moog, C., H.J. Fleury, I. Pellegrin, A. Kirn, and A.M. Aubertin, Autologous and heterologous neutralizing antibody responses following initial seroconversion in human immunodeficiency virus type 1-infected individuals. *J Virol*, 1997. **71**(5): p. 3734-3741.

93. Gray, E.S., P.L. Moore, I.A. Choge, J.M. Decker, F. Bibollet-Ruche, H. Li, . . . C.S. Team, Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. *J Virol*, 2007. **81**(12): p. 6187-6196.

94. Moore, P.L., N. Ranchobe, B.E. Lambson, E.S. Gray, E. Cave, M.-R. Abrahams, . . . N.C.f.H.A.V.I. (CHAVI), Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. *PLoS Pathog*, 2009. **5**(9): p. e1000598.

95. Moore, P.L., E.S. Gray, and L. Morris, Specificity of the autologous neutralizing antibody response. *Curr Opin HIV AIDS*, 2009. **4**(5): p. 358-363.

96. Piantadosi, A., D. Panteleeff, C.A. Blish, J.M. Baeten, W. Jaoko, R.S. McClelland, and J. Overbaugh, Breadth of neutralizing antibody response to human immunodeficiency virus type 1 is affected by factors early in infection but does not influence disease progression. *J Virol*, 2009. **83**(19): p. 10269-10274.

97. Gray, E.S., M.C. Madiga, T. Hermanus, P.L. Moore, C.K. Wibmer, N.L. Tumba, . . . C.S. Team, The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *J Virol*, 2011. **85**(10): p. 4828-4840.

98. Mikell, I., D.N. Sather, S.A. Kalams, M. Altfeld, G. Alter, and L. Stamatatos, Characteristics of the earliest cross-neutralizing antibody response to HIV-1. *PLoS Pathog*, 2011. **7**(1): p. e1001251.

99. Sanders, R.W., M. Venturi, L. Schiffner, R. Kalyanaraman, H. Katinger, K.O. Lloyd, . . . J.P. Moore, The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120. *J Virol*, 2002. **76**(14): p. 7293-7305.

100. Moore, J.P. and J. Sodroski, Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. *J Virol*, 1996. **70**(3): p. 1863-72.

101. Fouts, T.R., J.M. Binley, A. Trkola, J.E. Robinson, and J.P. Moore, Neutralization of the human immunodeficiency virus type 1 primary isolate JR-FL by human monoclonal antibodies correlates with antibody binding to the oligomeric form of the envelope glycoprotein complex. *J Virol*, 1997. **71**(4): p. 2779-2785.

102. Burton, D.R., J. Pyati, R. Koduri, S.J. Sharp, G.B. Thornton, P.W. Parren, . . . P.L. Nara, Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science*, 1994. **266**(5187): p. 1024-1027.

103. Wu, X., Z.-Y. Yang, Y. Li, C.-M. Hogerkorp, W.R. Schief, M.S. Seaman, . . . J.R. Mascola, Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science*, 2010. **329**(5993): p. 856-861.

104. Binley, J.M., T. Wrin, B. Korber, M.B. Zwick, M. Wang, C. Chappey, . . . D.R. Burton, Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *J Virol*, 2004. **78**(23): p. 13232-13252.

105. Wu, X., T. Zhou, S. O'Dell, R.T. Wyatt, P.D. Kwong, and J.R. Mascola, Mechanism of human immunodeficiency virus type 1 resistance to monoclonal antibody B12 that effectively targets the site of CD4 attachment. *J Virol*, 2009. **83**(21): p. 10892-10907.

106. Mao, Y., L. Wang, C. Gu, A. Herschhorn, A. Désormeaux, A. Finzi, . . . J.G. Sodroski, Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. *Proceedings of the National Academy of Sciences*, 2013. **110**(30): p. 12438-12443.

107. Scheid, J.F., H. Mouquet, B. Ueberheide, R. Diskin, F. Klein, T.Y.K. Oliveira, . . . M.C. Nussenzweig, Sequence and Structural Convergence of Broad and Potent HIV Antibodies That Mimic CD4 Binding. *Science*, 2011. **333**(6049): p. 1633-1637.

108. Zhou, T., I. Georgiev, X. Wu, Z.Y. Yang, K. Dai, A. Finzi, . . . P.D. Kwong, Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science*, 2010. **329**(5993): p. 811-7.

109. Li, Y., S. O'Dell, L.M. Walker, X. Wu, J. Guenaga, Y. Feng, . . . J.R. Mascola, Mechanism of neutralization by the broadly neutralizing HIV-1 monoclonal antibody VRC01. *J Virol*, 2011. **85**(17): p. 8954-8967.

110. Falkowska, E., A. Ramos, Y. Feng, T. Zhou, S. Moquin, L.M. Walker, . . . D.R. Burton, PGV04, an HIV-1 gp120 CD4 binding site antibody, is broad and potent in neutralization but does not induce conformational changes characteristic of CD4. *J Virol*, 2012. **86**(8): p. 4394-4403.

111. Rizzuto, C.D., R. Wyatt, N. Hernández-Ramos, Y. Sun, P.D. Kwong, W.A. Hendrickson, and J. Sodroski, A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science*, 1998. **280**(5371): p. 1949-1953.

112. Dorfman, T., M.J. Moore, A.C. Guth, H. Choe, and M. Farzan, A tyrosine-sulfated peptide derived from the heavy-chain CDR3 region of an HIV-1-neutralizing antibody binds gp120 and inhibits HIV-1 infection. *J Biol Chem*, 2006. **281**(39): p. 28529-35.

113. Moore, K.L., The biology and enzymology of protein tyrosine O-sulfation. *J Biol Chem*, 2003. **278**(27): p. 24243-24246.

114. Choe, H., W. Li, P.L. Wright, N. Vasilieva, M. Venturi, C.-c. Huang, . . . M. Farzan, Tyrosine sulfation of human antibodies contributes to recognition of the CCR5 binding region of HIV-1 gp120. *Cell*, 2003. **114**(2): p. 161-170.

115. Xiang, S.-H., M. Farzan, Z. Si, N. Madani, L. Wang, E. Rosenberg, . . . J. Sodroski, Functional mimicry of a human immunodeficiency virus type 1 coreceptor by a neutralizing monoclonal antibody. *J Virol*, 2005. **79**(10): p. 6068-6077.

116. Walker, L.M., M.D. Simek, F. Priddy, J.S. Gach, D. Wagner, M.B. Zwick, . . . D.R. Burton, A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals. *PLoS Pathog*, 2010. **6**(8): p. e1001028.

117. Walker, L.M., M. Huber, K.J. Doores, E. Falkowska, R. Pejchal, J.-P. Julien, . . . P. Poignard, Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature*, 2011. **477**(7365): p. 466-470.

118. Kitov, P.I. and D.R. Bundle, On the nature of the multivalency effect: a thermodynamic model. *J Am Chem Soc*, 2003. **125**(52): p. 16271-16284.

119. Zhu, X., C. Borchers, R.J. Bienstock, and K.B. Tomer, Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry*, 2000. **39**(37): p. 11194-11204.

120. Doores, K.J., C. Bonomelli, D.J. Harvey, S. Vasiljevic, R.A. Dwek, D.R. Burton, . . . C.N. Scanlan, Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proceedings of the National Academy of Sciences*, 2010. **107**(31): p. 13800-13805.

121. Trkola, A., M. Purtscher, T. Muster, C. Ballaun, A. Buchacher, N. Sullivan, . . . H. Katinger, Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol*, 1996. **70**(2): p. 1100-1108.

122. Calarese, D.A., C.N. Scanlan, M.B. Zwick, S. Deechongkit, Y. Mimura, R. Kunert, . . . I.A. Wilson, Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science*, 2003. **300**(5628): p. 2065-2071.

123. Platt, E.J., M.M. Gomes, and D. Kabat, Kinetic mechanism for HIV-1 neutralization by antibody 2G12 entails reversible glycan binding that slows cell entry. *Proceedings of the National Academy of Sciences*, 2012. **109**(20): p. 7829-7834.

124. McLellan, J.S., M. Pancera, C. Carrico, J. Gorman, J.-P. Julien, R. Khayat, . . . P.D. Kwong, Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature*, 2011. **480**(7377): p. 336-343.

125. Pancera, M., S. Shahzad-ul-Hussan, N.A. Doria-Rose, J.S. McLellan, R.T. Bailer, K. Dai, . . . P.D. Kwong, Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. *Nat. Struct. Mol. Biol.*, 2013. **20**(7): p. 804-813.

126. Julien, J.-P., J.H. Lee, A. Cupo, C.D. Murin, R. Derking, S. Hoffenberg, . . . A.B. Ward, Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9. *Proceedings of the National Academy of Sciences*, 2013. **110**(11): p. 4351-4356.

127. Löving, R., M. Sjöberg, S.-R. Wu, J.M. Binley, and H. Garoff, Inhibition of the HIV-1 spike by single-PG9/16-antibody binding suggests a coordinated-activation model for its three protomeric units. *J Virol*, 2013. **87**(12): p. 7000-7007.

128. Frey, G., H. Peng, S. Rits-Volloch, M. Morelli, Y. Cheng, and B. Chen, A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proceedings of the National Academy of Sciences*, 2008. **105**(10): p. 3739-3744.

129. Chakrabarti, B.K., L.M. Walker, J.F. Guenaga, A. Ghobbeh, P. Poignard, D.R. Burton, and R.T. Wyatt, Direct antibody access to the HIV-1 membrane-proximal external region positively correlates with neutralization sensitivity. *J Virol*, 2011. **85**(16): p. 8217-8226.

130. Cardoso, R.M.F., F.M. Brunel, S. Ferguson, M. Zwick, D.R. Burton, P.E. Dawson, and I.A. Wilson, Structural basis of enhanced binding of extended and helically constrained peptide epitopes of the broadly neutralizing HIV-1 antibody 4E10. *J Mol Biol*, 2007. **365**(5): p. 1533-1544.

131. Julien, J.-P., S. Bryson, J.L. Nieva, and E.F. Pai, Structural details of HIV-1 recognition by the broadly neutralizing monoclonal antibody 2F5: epitope conformation, antigen-recognition loop mobility, and anion-binding site. *J Mol Biol*, 2008. **384**(2): p. 377-392.

132. Huang, J., G. Ofek, L. Laub, M.K. Louder, N.A. Doria-Rose, N.S. Longo, . . . M. Connors, Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature*, 2012. **491**(7424): p. 406-412.

133. Haynes, B.F., J. Fleming, E.W. St Clair, H. Katinger, G. Stiegler, R. Kunert, . . . S.M. Alam, Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science*, 2005. **308**(5730): p. 1906-1908.

134. Verkoczy, L., M. Diaz, T.M. Holl, Y.-B. Ouyang, H. Bouton-Verville, S.M. Alam, . . . B.F. Haynes, Autoreactivity in an HIV-1 broadly reactive neutralizing antibody variable region heavy chain induces immunologic tolerance. *Proceedings of the National Academy of Sciences*, 2010. **107**(1): p. 181-186.

135. Brodsky, M.H., M. Warton, R.M. Myers, and D.R. Littman, Analysis of the site in CD4 that binds to the HIV envelope glycoprotein. *J Immunol*, 1990. **144**(8): p. 3078-3086.

136. Cochran, A.G., R.T. Tong, M.A. Starovasnik, E.J. Park, R.S. McDowell, J.E. Theaker, and N.J. Skelton, A minimal peptide scaffold for beta-turn display:

optimizing a strand position in disulfide-cyclized beta-hairpins. *J Am Chem Soc*, 2001. **123**(4): p. 625-632.

137. Chen, S., R.A. Chrusciel, H. Nakanishi, A. Raktabutr, M.E. Johnson, A. Sato, . . . M.I. Greene, Design and synthesis of a CD4 beta-turn mimetic that inhibits human immunodeficiency virus envelope glycoprotein gp120 binding and infection of human lymphocytes. *Proc Natl Acad Sci USA*, 1992. **89**(13): p. 5872-5876.

138. Vita, C., J. Vizzavona, E. Drakopoulou, S. Zinn-Justin, B. Gilquin, and A. Menez, Novel miniproteins engineered by the transfer of active sites to small natural scaffolds. *Biopolymers*, 1998. **47**(1): p. 93-100.

139. Kellenberger, E., G. Mer, C. Kellenberger, G. Marguerie, and J.F. Lefèvre, Solution structure of a conformationally constrained Arg-Gly-Asp-like motif inserted into the alpha/beta scaffold of leiurotoxin I. *Eur J Biochem*, 1999. **260**(3): p. 810-817.

140. Vita, C., C. Roumestand, F. Toma, and A. Ménez, Scorpion toxins as natural scaffolds for protein engineering. *Proc Natl Acad Sci USA*, 1995. **92**(14): p. 6404-6408.

141. Song, J., B. Gilquin, N. Jamin, E. Drakopoulou, M. Guenneugues, M. Dauplais, . . . A. Ménez, NMR solution structure of a two-disulfide derivative of charybdotoxin: structural evidence for conservation of scorpion toxin alpha/beta motif and its hydrophobic side chain packing. *Biochemistry*, 1997. **36**(13): p. 3760-3766.

142. Zinn-Justin, S., M. Guenneugues, E. Drakopoulou, B. Gilquin, C. Vita, and A. Ménez, Transfer of a beta-hairpin from the functional site of snake curaremimetic toxins to the alpha/beta scaffold of scorpion toxins: three-dimensional solution structure of the chimeric protein. *Biochemistry*, 1996. **35**(26): p. 8535-8543.

143. Vita, C., E. Drakopoulou, J. Vizzavona, S. Rochette, L. Martin, A. Ménez, . . . J.C. Gluckman, Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. *Proc Natl Acad Sci USA*, 1999. **96**(23): p. 13091-13096.

144. Zhang, W., G. Canziani, C. Plugariu, R. Wyatt, J. Sodroski, R. Sweet, . . . I. Chaiken, Conformational Changes of gp120 in Epitopes near the CCR5 Binding Site Are Induced by CD4 and a CD4 Miniprotein Mimetic †. *Biochemistry*, 1999. **38**(29): p. 9405-9416.

145. Martin, L., F. Stricher, D. Missé, F. Sironi, M. Pugnière, P. Barthe, . . . C. Vita, Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat Biotechnol*, 2003. **21**(1): p. 71-76.

146. Huang, C.-c., F. Stricher, L. Martin, J.M. Decker, S. Majeed, P. Barthe, . . . P.D. Kwong, Scorpion-toxin mimics of CD4 in complex with human immunodeficiency virus gp120 crystal structures, molecular mimicry, and neutralization breadth. *Structure*, 2005. **13**(5): p. 755-768.

147. Farzan, M., N. Vasilieva, C.E. Schnitzler, S. Chung, J. Robinson, N.P. Gerard, . . . J. Sodroski, A tyrosine-sulfated peptide based on the N terminus of CCR5 interacts with a CD4-enhanced epitope of the HIV-1 gp120 envelope glycoprotein and inhibits HIV-1 entry. *J Biol Chem*, 2000. **275**(43): p. 33516-33521.

148. Chiang, J.J., M.R. Gardner, B.D. Quinlan, T. Dorfman, H. Choe, and M. Farzan, Enhanced recognition and neutralization of HIV-1 by antibody-derived CCR5-mimetic peptide variants. *J Virol*, 2012. **86**(22): p. 12417-21.

149. Lu, M., S.C. Blacklow, and P.S. Kim, A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat. Struct. Mol. Biol.*, 1995. **2**(12): p. 1075-1082.

150. Tan, K., J. Liu, J. Wang, S. Shen, and M. Lu, Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc Natl Acad Sci U S A*, 1997. **94**(23): p. 12303-8.

151. Jiang, S., K. Lin, N. Strick, and A.R. Neurath, HIV-1 inhibition by a peptide. *Nature*, 1993. **365**(6442): p. 113.

152. Miyauchi, K., M.M. Kozlov, and G.B. Melikyan, Early steps of HIV-1 fusion define the sensitivity to inhibitory peptides that block 6-helix bundle formation. *PLoS Pathog*, 2009. **5**(9): p. e1000585.

153. Hines, J.C. and D.S. Ray, Construction and characterization of new coliphage M13 cloning vectors. *Gene*, 1980. **11**(3-4): p. 207-18.

154. Zacher, A.N., C.A. Stock, J.W. Golden, and G.P. Smith, A new filamentous phage cloning vector: fd-tet. *Gene*, 1980. **9**(1-2): p. 127-140.

155. Messing, J., B. Gronenborn, B. Muller-Hill, and P. Hans Hopschneider, Filamentous coliphage M13 as a cloning vehicle: insertion of a HindII fragment of the lac regulatory region in M13 replicative form in vitro. *Proc Natl Acad Sci U S A*, 1977. **74**(9): p. 3642-6.

156. Russel, M. and P. Model, Genetic analysis of the filamentous bacteriophage packaging signal and of the proteins that interact with it. *J Virol*, 1989. **63**(8): p. 3284-3295.

157. Lubkowski, J., F. Hennecke, A. Plückthun, and A. Wlodawer, Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure*, 1999. **7**(6): p. 711-722.

158. Smith, G.P., Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 1985. **228**(4705): p. 1315-1317.

159. McCafferty, J., A.D. Griffiths, G. Winter, and D.J. Chiswell, Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*, 1990. **348**(6301): p. 552-554.

160. Parmley, S.F. and G.P. Smith, Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene*, 1988. **73**(2): p. 305-318.

161. Bratkovic, T., Progress in phage display: evolution of the technique and its application. *Cell Mol Life Sci*, 2010. **67**(5): p. 749-67.

162. Bass, S., R. Greene, and J.A. Wells, Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins*, 1990. **8**(4): p. 309-314.

163. Qi, H., H. Lu, H.-J. Qiu, V. Petrenko, and A. Liu, Phagemid vectors for phage display: properties, characteristics and construction. *J Mol Biol*, 2012. **417**(3): p. 129-143.

164. Daugherty, P.S., Protein engineering with bacterial display. *Curr. Opin. Struct. Biol.*, 2007. **17**(4): p. 474-480.

165. Daugherty, P.S., G. Chen, M.J. Olsen, B.L. Iverson, and G. Georgiou, Antibody affinity maturation using bacterial surface display. *Protein Eng*, 1998. **11**(9): p. 825-832.

166. Georgiou, G., C. Stathopoulos, P.S. Daugherty, A.R. Nayak, B.L. Iverson, and R. Curtiss, Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat Biotechnol*, 1997. **15**(1): p. 29-34.

167. Feldhaus, M.J., R.W. Siegel, L.K. Opresko, J.R. Coleman, J.M.W. Feldhaus, Y.A. Yeung, . . . K.D. Wittrup, Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. *Nat Biotechnol*, 2003. **21**(2): p. 163-170.

168. van den Beucken, T., H. Pieters, M. Steukers, M. van der Vaart, R.C. Ladner, H.R. Hoogenboom, and S.E. Hufton, Affinity maturation of Fab antibody fragments by fluorescent-activated cell sorting of yeast-displayed libraries. *FEBS Lett*, 2003. **546**(2-3): p. 288-294.

169. Boder, E.T. and K.D. Wittrup, Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol*, 1997. **15**(6): p. 553-7.

170. Boder, E.T., K.S. Midelfort, and K.D. Wittrup, Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci USA*, 2000. **97**(20): p. 10701-5.

171. Blaise, L., A. Wehnert, M.P.G. Steukers, T. van den Beucken, H.R. Hoogenboom, and S.E. Hufton, Construction and diversification of yeast cell surface displayed

libraries by yeast mating: application to the affinity maturation of Fab antibody fragments. *Gene*, 2004. **342**(2): p. 211-218.

172. Weaver-Feldhaus, J.M., J. Lou, J.R. Coleman, R.W. Siegel, J.D. Marks, and M.J. Feldhaus, Yeast mating for combinatorial Fab library generation and surface display. *FEBS Lett*, 2004. **564**(1-2): p. 24-34.

173. Horlick, R.A., A.E. Schilling, P. Samama, R.N. Swanson, V.D. Fitzpatrick, A.K. Robbins, and B. Damaj, Combinatorial gene expression using multiple episomal vectors. *Gene*, 2000. **243**(1-2): p. 187-194.

174. Bowers, P.M., R.A. Horlick, T.Y. Neben, R.M. Toobian, G.L. Tomlinson, J.L. Dalton, . . . D.J. King, Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. *Proceedings of the National Academy of Sciences*, 2011. **108**(51): p. 20455-20460.

175. Ho, M. and I. Pastan, Display and selection of scFv antibodies on HEK-293T cells. *Methods Mol Biol*, 2009. **562**: p. 99-113.

176. Urban, J.H., R.M. Schneider, M. Compte, C. Finger, K. Cichutek, L. Alvarez-Vallina, and C.J. Buchholz, Selection of functional human antibodies from retroviral display libraries. *Nucleic Acids Res*, 2005. **33**(4): p. e35.

177. Taube, R., Q. Zhu, C. Xu, F. Diaz-Griffero, J. Sui, E. Kamau, . . . W.A. Marasco, Lentivirus display: stable expression of human antibodies on the surface of human cells and virus particles. *PLoS ONE*, 2008. **3**(9): p. e3181.

178. Horlick, R.A., J.L. Macomber, P.M. Bowers, T.Y. Neben, G.L. Tomlinson, I.P. Krapf, . . . D.J. King, Simultaneous surface display and secretion of proteins from mammalian cells facilitate efficient in vitro selection and maturation of antibodies. *J Biol Chem*, 2013. **288**(27): p. 19861-19869.

179. Cumbers, S.J., G.T. Williams, S.L. Davies, R.L. Grenfell, S. Takeda, F.D. Batista, . . . M.S. Neuberger, Generation and iterative affinity maturation of antibodies in vitro using hypermutating B-cell lines. *Nat Biotechnol*, 2002. **20**(11): p. 1129-1134.

180. Barbas, C.F., A.S. Kang, R.A. Lerner, and S.J. Benkovic, Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc Natl Acad Sci USA*, 1991. **88**(18): p. 7978-82.

181. Clackson, T., H.R. Hoogenboom, A.D. Griffiths, and G. Winter, Making antibody fragments using phage display libraries. *Nature*, 1991. **352**(6336): p. 624-628.

182. Kang, A.S., T.M. Jones, and D.R. Burton, Antibody redesign by chain shuffling from random combinatorial immunoglobulin libraries. *Proc Natl Acad Sci USA*, 1991. **88**(24): p. 11120-11123.

183. Waterhouse, P., A.D. Griffiths, K.S. Johnson, and G. Winter, Combinatorial infection and in vivorecombination: a strategy for making large phage antibody repertoires. *Nucleic Acids Res*, 1993. **21**(9): p. 2265-2266.

184. Sblattero, D. and A. Bradbury, Exploiting recombination in single bacteria to make large phage antibody libraries. *Nat Biotechnol*, 2000. **18**(1): p. 75-80.

185. Embleton, M.J., G. Gorochov, P.T. Jones, and G. Winter, In-cell PCR from mRNA: amplifying and linking the rearranged immunoglobulin heavy and light chain V-genes within single cells. *Nucleic Acids Res*, 1992. **20**(15): p. 3831-3837.

186. Low, N.M., P.H. Holliger, and G. Winter, Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. *J Mol Biol*, 1996. **260**(3): p. 359-368.

187. Irving, R.A., A.A. Kortt, and P.J. Hudson, Affinity maturation of recombinant antibodies using E. coli mutator cells. *Immunotechnology*, 1996. **2**(2): p. 127-143.

188. Martin, A. and F.X. Schmid, Evolutionary stabilization of the gene-3-protein of phage fd reveals the principles that govern the thermodynamic stability of two-domain proteins. *J Mol Biol*, 2003. **328**(4): p. 863-875.

189. Pham, P., R. Bransteitter, J. Petruska, and M.F. Goodman, Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*, 2003. **424**(6944): p. 103-107.

190. Goodman, M.F., M.D. Scharff, and F.E. Romesberg, AID-initiated purposeful mutations in immunoglobulin genes. *Adv. Immunol.*, 2007. **94**: p. 127-155.

191. Sidhu, S.S., H.B. Lowman, B.C. Cunningham, and J.A. Wells, Phage display for selection of novel binding peptides. *Meth Enzymol*, 2000. **328**: p. 333-363.

192. Dennis, M.S., M. Zhang, Y.G. Meng, M. Kadkhodayan, D. Kirchhofer, D. Combs, and L.A. Damico, Albumin binding as a general strategy for improving the pharmacokinetics of proteins. *J Biol Chem*, 2002. **277**(38): p. 35035-35043.

193. Li, B., H. Xi, L. Diehl, W.P. Lee, L. Sturgeon, J. Chinn, . . . S.S. Sidhu, Improving therapeutic efficacy of a complement receptor by structure-based affinity maturation. *J Biol Chem*, 2009. **284**(51): p. 35605-11.

194. Bahudhanapati, H., Y. Zhang, S.S. Sidhu, and K. Brew, Phage display of tissue inhibitor of metalloproteinases-2 (TIMP-2): identification of selective inhibitors of collagenase-1 (metalloproteinase 1 (MMP-1)). *J Biol Chem*, 2011. **286**(36): p. 31761-31770.

195. Hötzel, I., V. Chiang, J. Diao, H. Pantua, H.R. Maun, and S.B. Kapadia, Efficient production of antibodies against a mammalian integral membrane protein by phage display. *Protein Eng Des Sel*, 2011. **24**(9): p. 679-89.

196. Sui, J., W. Li, A. Murakami, A. Tamin, L.J. Matthews, S.K. Wong, . . . W.A. Marasco, Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc Natl Acad Sci USA*, 2004. **101**(8): p. 2536-41.

197. Valadon, P., J.D. Garnett, J.E. Testa, M. Bauerle, P. Oh, and J.E. Schnitzer, Screening phage display libraries for organ-specific vascular immunotargeting in vivo. *Proc Natl Acad Sci USA*, 2006. **103**(2): p. 407-412.

198. Mirzabekov, T., H. Kontos, M. Farzan, W. Marasco, and J. Sodroski, Paramagnetic proteoliposomes containing a pure, native, and oriented seven-transmembrane segment protein, CCR5. *Nat Biotechnol*, 2000. **18**(6): p. 649-54.

199. Smith, G.P. and J.K. Scott, Libraries of peptides and proteins displayed on filamentous phage. *Meth Enzymol*, 1993. **217**: p. 228-57.

200. Beerli, R.R., M. Bauer, R.B. Buser, M. Gwerder, S. Muntwiler, P. Maurer, . . . M.F. Bachmann, Isolation of human monoclonal antibodies by mammalian cell display. *Proc Natl Acad Sci USA*, 2008. **105**(38): p. 14336-41.

201. Ray, K., M.J. Embleton, B.L. Jailkhani, M.K. Bhan, and R. Kumar, Selection of single chain variable fragments (scFv) against the glycoprotein antigen of the rabies virus from a human synthetic scFv phage display library and their fusion with the Fc region of human IgG1. *Clin Exp Immunol*, 2001. **125**(1): p. 94-101.

202. Zwick, M.B., L.L. Bonnycastle, K.A. Noren, S. Venturini, E. Leong, C.F. Barbas, . . . J.K. Scott, The maltose-binding protein as a scaffold for monovalent display of peptides derived from phage libraries. *Anal Biochem*, 1998. **264**(1): p. 87-97.

203. Colwill, K., R.P.B.W. Group, and S. Gräslund, A roadmap to generate renewable protein binders to the human proteome. *Nat Methods*, 2011. **8**(7): p. 551-8.

204. Brunak, S., J. Engelbrecht, and S. Knudsen, Prediction of human mRNA donor and acceptor sites from the DNA sequence. *Journal of Molecular Biology*, 1991. **220**(1): p. 49-65.

205. Bendtsen, J.D., H. Nielsen, G. von Heijne, and S. Brunak, Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology*, 2004. **340**(4): p. 783-95.

206. Zettlmeissl, G., J.P. Gregersen, J.M. Duport, S. Mehdi, G. Reiner, and B. Seed, Expression and characterization of human CD4:immunoglobulin fusion proteins. *DNA Cell Biol*, 1990. **9**(5): p. 347-53.

207. Zhu, H. and R.A. Dean, A novel method for increasing the transformation efficiency of Escherichia coli-application forbacterial artificial chromosome library construction. *Nucleic Acids Res*, 1999. **27**(3): p. 910-1.

208. Rader, C. and C.F. Barbas, Phage display of combinatorial antibody libraries. *Curr Opin Biotechnol*, 1997. **8**(4): p. 503-8.

209. Li, M., F. Gao, J.R. Mascola, L. Stamatatos, V.R. Polonis, M. Koutsoukos, . . . D.C. Montefiori, Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol*, 2005. **79**(16): p. 10108-25.

210. Martin, L., F. Stricher, D. Misse, F. Sironi, M. Pugniere, P. Barthe, . . . C. Vita, Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat Biotechnol*, 2003. **21**(1): p. 71-6.

211. Diskin, R., J.F. Scheid, P.M. Marcovecchio, A.P. West, Jr., F. Klein, H. Gao, . . . P.J. Bjorkman, Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science*, 2011. **334**(6060): p. 1289-93.

212. Dalgleish, A.G., P.C. Beverley, P.R. Clapham, D.H. Crawford, M.F. Greaves, and R.A. Weiss, The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, 1984. **312**(5996): p. 763-7.

213. Quinlan, B.D., M.R. Gardner, V.R. Joshi, J.J. Chiang, and M. Farzan, Direct expression and validation of phage-selected peptide variants in mammalian cells. *J Biol Chem*, 2013.

214. Sullivan, N., Y. Sun, Q. Sattentau, M. Thali, D. Wu, G. Denisova, . . . J. Sodroski, CD4-Induced conformational changes in the human immunodeficiency virus type 1 gp120 glycoprotein: consequences for virus entry and neutralization. *J Virol*, 1998. **72**(6): p. 4694-703.

215. Seed, B. and A. Aruffo, Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc Natl Acad Sci USA*, 1987. **84**(10): p. 3365-3369.

216. Sullivan, N., Y. Sun, J. Li, W. Hofmann, and J. Sodroski, Replicative function and neutralization sensitivity of envelope glycoproteins from primary and T-cell line-passaged human immunodeficiency virus type 1 isolates. *J Virol*, 1995. **69**(7): p. 4413-22.

217. Choe, H. and M. Farzan, Chapter 7. Tyrosine sulfation of HIV-1 coreceptors and other chemokine receptors. *Methods Enzymol*, 2009. **461**: p. 147-70.

218. Choe, H. and M. Farzan, Tyrosine sulfate trapped by amber. *Nat Biotechnol*, 2006. **24**(11): p. 1361-2.

219. Liu, C.C. and P.G. Schultz, Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat Biotechnol*, 2006. **24**(11): p. 1436-40.

220. Murphy, G.J., G. Mostoslavsky, D.N. Kotton, and R.C. Mulligan, Exogenous control of mammalian gene expression via modulation of translational termination. *Nat Med*, 2006. **12**(9): p. 1093-1099.

221. Schenten, D., L. Marcon, G.B. Karlsson, C. Parolin, T. Kodama, N. Gerard, and J. Sodroski, Effects of soluble CD4 on simian immunodeficiency virus infection of CD4-positive and CD4-negative cells. *J Virol*, 1999. **73**(7): p. 5373-5380.

222. Haim, H., Z. Si, N. Madani, L. Wang, J.R. Courter, A. Princiotto, . . . J. Sodroski, Soluble CD4 and CD4-mimetic compounds inhibit HIV-1 infection by induction of a short-lived activated state. *PLoS Pathog*, 2009. **5**(4): p. e1000360.

223. Moore, J.P., J.A. McKeating, Y.X. Huang, A. Ashkenazi, and D.D. Ho, Virions of primary human immunodeficiency virus type 1 isolates resistant to soluble CD4 (sCD4) neutralization differ in sCD4 binding and glycoprotein gp120 retention from sCD4-sensitive isolates. *J Virol*, 1992. **66**(1): p. 235-243.

224. Désormeaux, A., M. Coutu, H. Medjahed, B. Pacheco, A. Herschhorn, C. Gu, . . . A. Finzi, The highly conserved layer-3 component of the HIV-1 gp120 inner domain is critical for CD4-required conformational transitions. *J Virol*, 2013. **87**(5): p. 2549-2562.

225. Cerutti, N., B.V. Mendelow, G.B. Napier, M.A. Papathanasopoulos, M. Killick, M. Khati, . . . A. Capovilla, Stabilization of HIV-1 gp120-CD4 receptor complex through targeted interchain disulfide exchange. *J Biol Chem*, 2010. **285**(33): p. 25743-25752.

226. Martin, G., B. Burke, R. Thaï, A.K. Dey, O. Combes, O.H.P. Ramos, . . . L. Martin, Stabilization of HIV-1 envelope in the CD4-bound conformation through specific cross-linking of a CD4 mimetic. *J Biol Chem*, 2011. **286**(24): p. 21706-21716.

227. Liu, C.C., A.V. Mack, M.-L. Tsao, J.H. Mills, H.S. Lee, H. Choe, . . . V.V. Smider, Protein evolution with an expanded genetic code. *Proc Natl Acad Sci USA*, 2008. **105**(46): p. 17688-17693.

228. Mahalanabis, M., P. Jayaraman, T. Miura, F. Pereyra, E.M. Chester, B. Richardson, . . . N.L. Haigwood, Continuous Viral Escape and Selection by Autologous Neutralizing Antibodies in Drug-Naive Human Immunodeficiency Virus Controllers. *J Virol*, 2008. **83**(2): p. 662-672.

229. Sather, D.N., S. Carbonetti, J. Kehayia, Z. Kraft, I. Mikell, J.F. Scheid, . . . L. Stamatatos, Broadly neutralizing antibodies developed by an HIV-positive elite neutralizer exact a replication fitness cost on the contemporaneous virus. *J Virol*, 2012. **86**(23): p. 12676-12685.

230. Diskin, R., F. Klein, J.A. Horwitz, A. Halper-Stromberg, D.N. Sather, P.M. Marcovecchio, . . . P.J. Bjorkman, Restricting HIV-1 pathways for escape using rationally designed anti-HIV-1 antibodies. *J Exp Med*, 2013. **210**(6): p. 1235-49.

231. Orloff, S.L., M.S. Kennedy, A.A. Belperron, P.J. Maddon, and J.S. McDougal, Two mechanisms of soluble CD4 (sCD4)-mediated inhibition of human immunodeficiency virus type 1 (HIV-1) infectivity and their relation to primary HIV-1 isolates with reduced sensitivity to sCD4. *J Virol*, 1993. **67**(3): p. 1461-1471.

232. Zhang, Y.J., R. Fredriksson, J.A. McKeating, and E.M. Fenyö, Passage of HIV-1 molecular clones into different cell lines confers differential sensitivity to neutralization. *Virology*, 1997. **238**(2): p. 254-264.

233. Mangeat, B., P. Turelli, G. Caron, M. Friedli, L. Perrin, and D. Trono, Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, 2003. **424**(6944): p. 99-103.

234. van Gils, M.J., Z. Euler, B. Schweighardt, T. Wrin, and H. Schuitemaker, Prevalence of cross-reactive HIV-1-neutralizing activity in HIV-1-infected patients with rapid or slow disease progression. *AIDS*, 2009. **23**(18): p. 2405-2414.

235. Doria-Rose, N.A., R.M. Klein, M.G. Daniels, S. O'Dell, M. Nason, A. Lapedes, . . . M. Connors, Breadth of human immunodeficiency virus-specific neutralizing activity in sera: clustering analysis and association with clinical variables. *J Virol*, 2010. **84**(3): p. 1631-1636.